Educational Technology & Society
An International Journal

Aims and Scope

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

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

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

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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


Founding Editor
Kinsshuk, Athabasca University, Canada.

Editors

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

Advisory Board

Ignacio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jemni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasio 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 Lelouch, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McGreal, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö University, Sweden; Riichiro Mizoguchi, Osaka University, Japan; Permannad 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; Slavík Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-Chung Tsaï, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan; Yu-Mei Wang, University of Alabama at Birmingham, USA; Ashok Patel, CAL Research & Software Engineering Centre, UK; Reinhard Oppermann, Fraunhofer Institut Angewandte Informationstechnik, Germany; Vladimir A Fomichev, K. E. Tsolikovsky Russian State Tech Univ, Russia; Olga S Fomicheva, Studio “Culture, Ecology, and Foreign Languages,” Russia; Piet Kommers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; Erkki Sutinen, University of Joensuu, Finland; Vladimir Uskov, Bradley University, USA.

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

Technical Manager
Panagiotis Zervas, University of Piraeus & Centre for Research and Technology Hellas, Greece

Executive Peer-Reviewers

http://www.ifets.info/
Supporting Organizations
National Sun Yat-sen University, Taiwan
Centre for Research and Technology Hellas, Greece
Athabasca University, Canada

Subscription Prices and Ordering Information
For subscription information, please contact the editors at ets-editors@ifets.info.

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

Abstracting and Indexing

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

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

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

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

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

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

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

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

### Guest Editorial

Intelligent and Affective Learning Environments: New Trends and Challenges  
*Ramón Zatarain-Cabada, Giner Alor-Hernández, María Lucía Barrón-Estrada, Ricardo Colomo-Palacios and Hao-Chiang Koong Lin*  
1–4

### Special Issue Articles

#### PlayIT: Game Based Learning Approach for Teaching Programming Concepts

*Anuradha Mathrani, Shelly Christian and Agate Ponder-Sutton*  
5–17

#### KAPEAN: Understanding Affective States of Children with ADHD

*Fernando Martínez, Claudia Barreza, Nimrod González and Juan González*  
18–28

#### Affective Behavior and Nonverbal Interaction in Collaborative Virtual Environments

*Adriana Peña, Nora Rangel, Mirna Muñoz, Jezreel Mejia and Graciela Lara*  
29–41

#### A Computational Model of Learners Achievement Emotions Using Control-Value Theory

*Karla Muñoz, Julieta Noguez, Luis Neri, Paul Mc Kevitt and Tom Lunney*  
42–56

#### Cross Space: The Exploration of SNS-Based Writing Activities in a Multimodal Learning Environment

*Kwang-Soon Lee and Bong-Gyu Kim*  
57–76

#### Identification of Action Units Related to Affective States in a Tutoring System for Mathematics

*Gustavo Padrón-Rivera, Genaro Rebolledo-Mendez, Pilar Pozos Parra and N. Sofia Huerta-Pacheco*  
77–86

#### Analyzing How Emotion Awareness Influences Students’ Motivation, Engagement, Self-Regulation and Learning Outcome

*Marta Arguedas, Thanasis Daradoumis and Fatos Xhafa*  
87–103

#### The Effects of Positive and Negative Mood on Cognition and Motivation in Multimedia Learning Environment

*Tze Wei Liew and Su-Mae Tan*  
104–115

#### Designing Empathetic Animated Agents for a B-Learning Training Environment within the Electrical Domain

*Yasmin Hernández, Miguel Pérez-Ramírez, Ramón Zatarain-Cabada, Lucía Barrón-Estrada and Giner Alor-Hernández*  
116–131

#### Construction of Multi-mode Affective Learning System: Taking Affective Design as an Example

*Hao-Chiang Koong Lin, Sheng-Hsiung Su, Ching-Ju Chao, Cheng-Yen Hsieh and Shang-Chin Tsai*  
132–147

### Full Length Articles

#### A Review on the Use of Robots in Education and Young Children

*Lai Poh Emily Toh, Albert Causo, Pei-Wen Tzuo, I-Ming Chen and Song Huat Yeo*  
148–163

#### Tabletops for Peace: Technology Enhanced Peacemaking in School Contexts

*Andri Ioannou and Chrystalla Antoniou*  
164–176

#### Exploring Engaging Gamification Mechanics in Massive Online Open Courses

*Jen-Wei Chang and Hung-Yu Wei*  
177–203

#### Subgroup Discovery with User Interaction Data: An Empirically Guided Approach to Improving Intelligent Tutoring Systems

*Eric G. Poitras, Susanne P. Lajoie, Tenzin Doleck and Amanda Jarrell*  
204–214

#### Toward A Unified Modeling of Learner’s Growth Process and Flow Theory

*Geiser C. Challco, Fernando R. H. Andrade, Simone S. Borges, Iq I. Bittencourt and Seiji Isotani*  
215–227
Identifying Effective Design Features of Technology-Infused Inquiry Learning Modules: A Two-Year Study of Students’ Inquiry Abilities
Ying-Shao Hsu, Su-Chi Fang, Wen-Xin Zhang, Hsin-Kai Wu, Pai-Hsing Wu and Fu-Kwun Hwang
228–244

A Social Learning Management System Supporting Feedback for Incorrect Answers based on Social Network Services
Jiseong Son, Jeong-Dong Kim, Hong-Seok Na and Doo-Kwon Baik
245–257

Group Formation in Mobile Computer Supported Collaborative Learning Contexts: A Systematic Literature Review
Sofiane Amara, Joaquim Macedo, Fatima Bendella and Alexandre Santos
258–273

The Search for Pedagogical Dynamism – Design Patterns and the Unselfconscious Process
Agata Mouasher and Jason M. Lodge
274–285

The Impact of Peer Review on Creative Self-efficacy and Learning Performance in Web 2.0 Learning Activities
Chen-Chung Liu, Kuan-Hsien Lu, Leon Yufeng Wu and Chin-Chung Tsai
286–297

Academic Achievements and Satisfaction of the Clicker-Aided Flipped Business English Writing Class
Yu Zhonggen and Wang Guifang
298–312

A Review of Research on Intercultural Learning through Computer-Based Digital Technologies
Emrullah Yasin Çiftçi
313–327

Podcasts: A Factor to Improve Iranian EFL Learner’ Self-Regulation Ability and Use of Technology
Somayeh Naseri and Khalil Motallebzadeh
328–339

Curriculum Integration of MALL in L1/L2 Pedagogy: Perspectives on Research
Shu-Mei Gloria Chwo, Michael W. Marek and Wen-Chi Vivian Wu
340–354

Mining Learning and Crafting Scientific Experiments: A Literature Review on the Use of Minecraft in Education and Research
Steve Nebel, Sascha Schneider and Günter Daniel Rey
355–366

Role-Play Game-Enhanced English for a Specific-Purpose Vocabulary-Acquisition Framework
Fang-Chen Lu and Ben Chang
367–377

Enacting Viewing Skills with Apps to Promote Collaborative Mathematics Learning
Kay Yong, Khoo
378–390

Erratum

Erratum to - Volume 17, Issue 4, pp. 229-241
391
Guest Editorial: Intelligent and Affective Learning Environments: New Trends and Challenges

Ramón Zatarain-Cabada¹, Giner Alor-Hernández², María Lucía Barrón-Estrada¹, Ricardo Colomo-Palacios³ and Hao-Chiang Koong Lin⁴

¹Instituto Tecnológico de Culiacán, México // ²Instituto Tecnológico de Orizaba, México // ³Ostfold University College, Norway // ⁴National Taipei University of Business, Taiwan // rzatarain@itculiacan.edu.mx, lbarron@itculiacan.edu.mx // galor@itorizaba.edu.mx // ricardo.colomo-palacios@hibo.no // koonglin@gmail.com

Technology poses a huge potential in the educational field (Stantchev et al., 2014). As a result of this, researchers in this field of study devote great part of their efforts on finding better technological solutions (Vásquez-Ramírez et al., 2014). Within this field, Traditional Intelligent Tutoring Systems (ITS) are able to support and control students’ learning at several levels; however, it does not provide space for student-driven learning and knowledge acquisition. From this perspective, Intelligent Learning Environments and similar tutoring systems have emerged as a type of intelligent educational system that combines the features of traditional ITS with learning environments. This kind of educational system can be very helpful in supporting human learning by using Artificial Intelligence (AI) techniques, transforming information into knowledge, using it for tailoring many aspects of the educational process to the particular needs of each actor, and timely providing useful suggestions and recommendations (Brusilovsky et al., 1993; Carbonell, 1970; Clancey, 1979; Anderson et al., 1990; Aleven & Koedinger, 2002; Woolf, 2009).

In addition to traditional cognitive state identification, ITS have recently incorporated the ability to recognize the emotions of students (Calvo & D’Mello, 2010; Wolf et al., 2009; Baker et al., 2010). These tutoring systems can detect the affective states of learners by using different types of data sources such as dialogs, speech, physiology, and facial expressions (Zeng et al., 2009; Calvo & D’Mello, 2010; Arroyo et al., 2009; Conati & Maclaren, 2009; Burleson, 2011). Moreover, they seek to transform negative states of students (e.g., confusion) into positive (e.g., commitment) in order to facilitate appropriate emotional conditions for learning. Affective Tutoring Systems identify confusion, frustration, boredom, engagement, and other prominent emotions during learning activities (D’Mello & Graesser, 2012; D’Mello et al., 2014; Graesser & D’Mello, 2012). The recognition of students’ affective states can be implemented by different machine learning techniques, such as Bayesian Networks (Conati & Maclaren, 2009), Hidden-Markov Models (D’Mello & Graesser, 2010), or Neural Networks (Moridis & Economides, 2009). Although many works and studies have considered the development of affective tutoring systems, no research works have yet focused on Intelligent and Affective Learning Environments, where components involved in the environment (the learning environment, the intelligent tutoring system, and/or the adaptive system) support the learning process. Therefore, it is necessary to propose new approaches, techniques, methods, and processes in the field of Intelligent and Affective Learning Environments in order to consider cognitive and affective aspects in the teaching-learning and decision making processes.

This special issue of Journal of Educational Technology & Society (ET&S) on Intelligent and Affective Learning Environments: New Trends and Challenges, contains one kind of contribution: regular research papers. These works have been edited according to the norms and guidelines of JETS. Several call for papers were distributed among the main mailing lists of the field for researchers to submit their works to this issue. In the first deadline, we received a total of 32 expressions of interest in the form of abstracts. Due to the large amount of submissions, abstracts were subject to a screening process to ensure their clarity, authenticity, and relevancy to this special issue. Proposals came from several countries such as Algeria, Bosnia and Herzegovina, Brazil, Canada, Colombia, Denmark, Germany, Greece, India, Ireland, the Republic of Korea, Malaysia, Malta, Mexico, New Zealand, Norway, Philippines, Poland, Romania, Serbia, Spain, Taiwan, Tunisia, Turkey, United Kingdom of Great Britain, Northern Ireland, and United States of America.

After the screening process, 25 papers were invited to submit full versions. At least two reviewers were assigned to every work to proceed with the peer review process. Ten papers were finally accepted for their publication after corrections requested by reviewers and editors were addressed. The ten regular research papers introduce novel and interesting results in the form of theoretical and experimental research and case studies that apply new perspectives on Intelligent and Affective Learning Environments.
The special issue opens with a research paper proposed by Shelly Christian and Anuradha Mathrani. It proposes a game-based learning (GBL) approach to engage students in learning and enhance their programming skills. The work explains in detail how an educational game was mapped with the curriculum of a prescribed programming course in a computing study program. The main contribution of this work is the use of gaming elements for ongoing development of innovative pedagogies in teaching and learning.

Also, Fernando Martínez Reyes et al. describes a case based study carried out to observe and understand the level of attention, cognition, and memory of children with ADHD (Attention Deficit Hyperactivity Disorder). In this third research, the author developed a system called KAPEAN that collects information about the affective states that are present during the educational game, as it aims to understand how children’s emotions should be considered for future designs of adaptive tutoring tools.

Adriana Peña et al. introduce the third paper of the special issue. The research proposes the analysis of nonverbal interaction to identify affective behavior during the accomplishment of collaborative tasks, which can be extended to verbal content. An exploratory study was conducted to understand the possibilities of this approach. Results showed behavior patterns that can be associated with task-focused affective states.

The fourth paper by Karla Muñoz et al. describes the design and evaluation of an emotional student model of Control-value theory applied to online Game-Based Learning environments using Approach 2. The model was implemented by using a dynamic sequence of Bayesian Networks (BNs). PlayPhysics - an emotional GBL environment for teaching Physics - was designed, implemented, and evaluated. Cross-validation and Cohen’s Kappa were used to evaluate the model.

The fifth paper, proposed by Kwang-Soon Lee and Bong-Gyu Kim, explores the positive learning effect of formulating English sentences via Social Network Service (SNS; Kakao-Talk). Results support the benefit of correlated factors: the efficiency of input, challenge, social presence, the usefulness of to-verb practice, and feedback. The research work emphasizes on the potential of using SNS as an educational platform and highlights the necessity of a transactional space that combines classroom-based education with enjoyable leisure activities for young people.

Gustavo Padrón-Rivera et al. presents an analysis regarding the identification of action units (AUs) related to affective states and their impact on learning with a tutoring system. To assess affective states, a tool was devised to identify AUs on pictures of human faces. The paper aims to show that AUs are easier means to analyze affective states.

In the following paper, Marta Arguedas, Thanasis Daradoumis and FatosXhafa analyze the effects of emotion awareness, supported by specific teaching strategies, on students’ motivation, engagement, self-regulation and learning outcome in long-term blended collaborative learning practices. Results show that when students are aware of their emotions and guided by specific teaching strategies, their learning performance improves in relation to their motivation, engagement, and self-regulation. Likewise, findings suggest that when teachers are conscious of students' emotional states, their attitudes and feedback become more effective and timely.

The eighth paper by Tze Wei Liew and Su-Mae Tan presents a study of the effects of positive and negative mood on cognition and motivation in multimedia learning environments. The work is supported by two experiments. Experiment 1 was conducted to determine whether facilitating effects of positive mood found in previous works could be replicated with a multimedia learning system that teaches basic programming algorithm to learners at an Asian university. Experiment 2 researched the effects of induced negative mood on learners’ cognition and motivation in a multimedia learning environment.

Yasmin Hernandez et al. proposes a b-learning model to support adaptive and distance training for electrician examinations. The adaptation is based on a representation of the trainees’ knowledge and affective states. An animated pedagogical agent was developed for this study. The agent guides trainees and provides them with instructions. It also deploys different facial expressions that convey emotions and empathy to trainees. These facial expressions were incorporated into the learning model in order to achieve believability, social interaction, and user engagement, and hence improve learning.
The last paper, proposed by Hao-Chiang Koong Lin et al., studies the design of a non-simultaneous distance education system with affective computing. Such system integrates interactive agent technology with the curricular instruction of affective design. The curricular collocation system design used affective design as the learning material to provide respondents with more impressive learning experiences. Prototype assessment and final assessment were employed to discuss the usability of the system and its satisfaction among users.

Once a brief summary of papers has been provided, we would also like to express our gratitude to the reviewers who kindly accepted to contribute in the evaluation of papers at all stages of the editing process. We equally and especially wish to thank the Editor-in-Chief, Demetrios G. Sampson, for grating us the opportunity to edit this special issue and providing valuable comments to improve the selection of research works.

**Reviewer board**

Adriana Peña, Universidad de Guadalajara, México
Alejandro Rodríguez-Gonzalez, Polytechnic University of Madrid, Spain
Carlos A. Reyes-García, INAOE, México
Chih-yung Chen, St. John's University, Taiwan
Claude Frasson, University of Montreal, Canada
Cristina Casado Lumbreras, Universidad Internacional de La Rioja, Spain
Fazel Keshtkar, Southeast Missouri State University, USA
Genaro Rebolledo-Méndez, Universidad Veracruzana, México
Giner Alor-Hernández, Instituto Tecnológico de Orizaba, México
Gonzalo Méndez, Universidad Complutense de Madrid, Spain
Hao-Chiang Koong Lin, National Taipei University of Business, Taiwan
Harald Holone, Østfold University College, Norway
Jaime Alberto Guzmán-Luna, National University of Colombia, Colombia
Javier González Sánchez, Arizona State University, USA
José Luis Sánchez Cervantes, Instituto Tecnológico de Orizaba, México
José-Maria Álvarez-Rodríguez, Universidad Carlos III de Madrid, Spain
Juan Miguel Gómez Berbis, Universidad Carlos III de Madrid, Spain
Juliesta Noguex Monroy, ITESM, México
Lisbeth Rodriguez Mazahua, Instituto Tecnológico de Orizaba, México
María Elena Chávez Echeagaray, Arizona State University, USA
María Lucía Barrón-Estrada, Instituto Tecnológico de Culiacán, México
María Mercedes T. Rodrigo, Ateneo de Manila University, Philippines
Mario Arturo Gutiérrez Alonso, Bread and Butter SA, Switzerland
Miguel Ángel Conde, Universidad de León, Spain
Miguel Pérez-Ramírez, Instituto de Investigaciones Eléctricas, México
Mihai Dascălu, University Politehnica of Bucharest, Romania
Miroslav Minović, University of Belgrade, Serbia
Patricia Ordoñez de Pablos, Oviedo University, Spain
Radoslaw Michalski, Wroclaw University of Technology, Poland
Rafael Valencia-García, Murcia University, Spain
Ramón Zatarain-Cabada, Instituto Tecnológico de Culiacán, México
Ricardo Colomo-Palacios, Østfold University College, Norway
Robin Wolff, German Aerospace Center (DLR), Germany
Rodney D. Nielsen, University of North Texas, USA
Roger Azevedo, North Carolina State University, USA
Roger Nkambou, University of Quebec at Montreal, Canada
Sanjay Misra, Atılım University, Turkey
Vladimir Stantchev, SRH University Berlin, Germany
Yasmin Hernández-Pérez, Instituto de Investigaciones Eléctricas, México
References


PlayIT: Game Based Learning Approach for Teaching Programming Concepts

Anuradha Mathrani*, Shelly Christian and Agate Ponder-Sutton
School of Engineering and Advanced Technology, Massey University, Auckland, New Zealand // a.s.mathrani@massey.ac.nz // shelly.christian@outlook.com // a.m.ponder-sutton@massey.ac.nz

*Corresponding author

ABSTRACT
This study demonstrates a game-based learning (GBL) approach to engage students in learning and enhance their programming skills. The paper gives a detailed narrative of how an educational game was mapped with the curriculum of a prescribed programming module in a computing diploma study programme. Two separate student cohorts were invited to participate in the GBL experiment. One student cohort had not yet started study of the programming module, while the second student cohort had recently completed the module on programming. Findings reveal that educational games add to the fun element in learning, and students rated the game as an effective way to learn programming. Students could easily relate gaming elements to difficult programming constructs. Animated game scenarios showed high levels of engagement among students. Some students found the use of gaming elements as a better way to express their program’s logic when giving oral presentations for the final assessment. Results indicate that a majority of study participants passed the programming module in the first attempt. The study contributes to the use of gaming elements for ongoing development of innovative pedagogies in teaching and learning.

Keywords
Game based learning, ICT education, Programming, Educational games, Teaching and learning pedagogies

Introduction
The application of technology-enabled solutions in everyday activities has a pervasive effect on information and communications technology (ICT) education. There is an increasing demand for the analytical, technical and programming abilities of information technology (IT) graduates by the computing industry. To build the problem solving capabilities in students, ICT courses are designed with many practical elements. However, after entering into ICT related courses (e.g., programming, networks and databases), many students find it difficult to transmit taught concepts to real world applications. These students may find courses to be dry and boring, which lowers their motivation and interest in learning (Prensky, 2003; Sarkar, 2006). If students are not interested or motivated, it is difficult to keep them engaged in classrooms. To enhance student learning for achieving required IT based skill sets, innovative pedagogical approaches are applied to teaching and learning (T&L) practices. Behavioural scientists suggest utilising fun based interventions to engage active learning as an effective pedagogical approach (Dicheva et al., 2015; Oblinger, 2006). One such approach for adding engaging elements to classrooms is use of game-based learning (GBL) or serious games, whereby people of all ages and genders can play games for many hours without realising they are potentially in a T&L environment (Soflano, 2011). Gaming activities are a good source of engagement and bring fun into learning by providing instant gratification to players when tasks are completed successfully, allowing them to reach higher stages in the gameplay. Many workplaces use gamification strategies to empower employees, with one German automotive company Volkswagen labelling gameplay strategy as “the fun theory” (Huang & Soman, 2013).

Directed instruction is a venerable part of the classroom environment; lectures are designed to explain theoretical concepts, which are complemented with practical experiments. Teachers evaluate student learning with a set of formal assignments, oral presentations and written exams. Nevertheless, T&L environments could be made more fun if critical skills are taught both through directed teaching and game-mediated interventions. This would make students more engaged and motivated, and could change the student’s mindset that the journey of learning is not dry or boring, but can be fun. This study attempts to address this gap by utilising a GBL approach to T&L context using a case involved in ICT education.

The paper first gives some highlights of current literature in pedagogical approaches to ICT education, and how educational games have been used in previous studies. The case study of PlayIT (a pseudonym) is introduced next, followed by an explanation of how the chosen educational game has been mapped with the subject module for programming in an ICT course curriculum. The design of a GBL experiment with two different student cohorts is presented. The paper then discusses experiment findings using quantitative and qualitative methods to identify any
significant/insignificant correlations with diverse student cohort datasets. Student results are further investigated to inform how study participants progressed in their subject knowledge. Finally, the paper concludes with an overview of our findings, leading to further contributions in the ongoing quest for innovative, useful pedagogies in T&L environments.

**Pedagogical approaches to ICT education**

Computing is interwoven in almost all facets of managing and running a business. Furthermore, it is expected that technological applications will get more efficient and advanced over time, requiring more skilled and collaborative workforce (Stantchev, Prieto-González, & Tammi, 2015). A study investigating critical information systems/information technology (IS/IT) skills from the perspectives of seventy managers shows that web applications, online services, networking protocols, wireless communications and wireless applications are the skills of the future (Lee & Mirchandani, 2010). Moreover, growing use of technology in our daily lives has added to the myriad of technology courses offered by education providers to prepare upcoming students. ICT education as such provides “an effective link between purpose, people and pedagogy inside the institutions” (Stensaker et al., 2007, p. 431). Students enrol in ICT courses to learn new technologies and to comprehend the bigger picture of how IT solutions are being developed for businesses. However, these students face many challenges in grasping conceptual understanding and logical reasoning of how classroom topics in hardware, programming, databases or networks are related with real world applications. “Students could not transfer knowledge gained from either lectures or theoretical exercises to practical exercises. Without having direct hardware interaction, students learning becomes abstract, which leads to their displeasure and to the main question: Why we are learning this, and how and where shall I use it?” (Stolikj, Ristov & Ackovska, 2011, p. 340). Students’ acceptance of technology has been shown to be a critical factor in understanding by Stantchev et al. (2014). To help students relate course contents to real world examples, a blended learning approach has been used. This approach consists of a mix of teaching deliveries, namely (1) Classroom: traditional teaching, (2) Website: web based self-paced learning, (3) Actual lab: real experiments, and (4) Virtual lab: visualization/animation techniques (Xie, Li & Geng, 2008).

Studies in ICT education suggest that students find it challenging to apply taught concepts to a problem when there is no single, simple or well-known solution. “Students can also display an inability to translate classroom examples to other domains with analogous scenarios, betraying a lack of analytical problem-solving skills. For the students, these problems can lead to confusion, a lack of self-confidence and a lack of motivation to continue” (Connolly & Stansfield, 2006, p. 462). To overcome these challenges, it is suggested that classroom teaching be scaffolded with interactive computer games to simulate problem-based scenarios, since games provide more opportunities for collaboration and reflection, which in turn will lead to increased motivation (Connolly, Stansfield, & McLellan, 2006). Papastergiou (2009) evaluated the learning effectiveness and motivational appeal of a computer game targeted at the learning of computer memory concepts for high school students. Results showed the gaming approach to be very effective in gaining students’ understanding of computer memory concepts. Papastergiou concludes not only the learning effectiveness, but, also provides solution to the students “feeling bored.” One participant in Papastergiou’s study responded: “It’s more enjoyable and active. You never get bored as in traditional teaching because you concentrate on a goal.” Ebner and Holzinger (2007) used an educational game IFM (Internal Force Master) in a mechanical engineering study programme. Their findings demonstrated high levels of user empowerment and fun elements for students who played IFM. The feedback in the Ebner and Holzinger study showed students’ readiness to play the game a second time in the event of a failure. However, the study did not find noticeable difference in students’ results between those students whose learning involved IFM game play, and those students who had learned in a traditional classroom environment.

Another study has described design of an educational framework through an iterative development process resulting in a game that can improve student engagement, satisfaction and skills transfer (Barnes et al., 2007). Researchers used students who had completed at least one computer science course and were moderately familiar with programming concepts. Each participant was given two games to play, that is, “Saving Princess Sera” and “The Catacombs.” Participants gave pre-test prior to the game play and post-test after the game play. Both the tests contained problems where students had to determine the outcome of programming components such as “if-then-else” statements and “while” loops. Study found that despite poor test results students were able to understand most of the programming quests and feedback of students was extremely positive. Computer games are thus transformed into
social experiences within classroom settings, to offer a constructivist approach that are interactive in nature and help to generate meaning in learning (Hämäläinen, 2011).

The case study design

The case study described here as PlayIT is a non-university education provider offering ICT related subject courses at different study levels designed by New Zealand Qualification Authority (NZQA). NZQA is a government organisation responsible for managing the New Zealand Qualifications Framework (NZQF), and is the source for accurate and current information on quality assured qualifications in New Zealand. NZQA administers secondary school assessment system, provides independent quality assurance of non-university education providers, and recognizes qualifications by setting specific achievement and unit standards for approved courses (http://www.nzqa.govt.nz). PlayIT runs three computing courses, namely, National Diploma in Computing (level 5), Diploma in Computer Networking and Security (level 6), and Diploma in Networks and Security (level 7).

![Course structure at PlayIT](image)

Only one of these courses includes programming; which is the “National Diploma in Computing” (NDC) at level 5. The NDC contains five curriculum modules: databases, hardware, networking, software engineering and programming. Each module within NDC contains one or more unit standards (US). The programming module consists of two US, namely US-6774 (a basic level of programming using a procedural or event driven approach) which is a prerequisite of US-6776 (an advanced level of programming using an object-oriented approach). Figure 1 shows the detail of the course structure at PlayIT for NDC curriculum.

In March 2014, interviews were conducted with experienced IT tutors at PlayIT to identify learning challenges faced by students pursuing ICT curriculum. Interview findings revealed tutor’s perceptions of issues faced by students and suggestions to mitigate those issues. The identified issues are: difficulty in transferring theoretical knowledge to practical exercise, difficulty in relating course contents to industry use, lack of interest in course contents, and difficulty in understanding conceptual topics due to lack of analytical and logical skills. Tutors suggestions likewise affirmed the use of GBL strategy to encourage participation and bring about active learning through simulations of problem based scenarios using some animated environment. Furthermore, tutors said topics related to programming constructs were challenging for students. Accordingly, US-6774 from the subject module for programming from the NDC course was selected as the module where the GBL experiment would be applied. Programming is broad topic
and contains many sub-topics, however, the fundamental components; such as sequential logic flow, if-then-else, loops, functions and recursions; were selected for the GBL experiment.

PlayIT runs many courses in parallel with different student cohorts. For this study, two student cohorts were selected who were at different stages in their level 5 study programme. Cohort 1 comprised of 20 students who had not yet started study of the US-6774 module, while cohort 2 students had 24 students who had recently completed the US-6774 module, but had not yet been assessed. An educational game (LightBot 2.0) was selected for investigating the effectiveness of GBL in T&L environments. An experiment utilising this game was conducted in May 2014 with both student cohorts in two separate classroom settings. Student feedback was collected in two stages, first immediately after the gameplay, and second time in September 2014 after all students had completed final assessments of the programming module.

Figure 2. Case study design

Figure 2 shows the research design for the case study. The design consists of six processes: (1) establish criteria, (2) selection of study participants, (3) problem definition, (4) treatment or intervention strategy, (5) techniques and methods, and (6) analysis and results.

Mapping of educational game with programming module

Two considerations in selecting the educational game were its relevance to curriculum topics and coverage of minimum number of topics. Many games were investigated to confirm alignment of game elements with learning activities outlined in the programming course module of the NDC. After detailed search of ICT educational games, the game Light Bot 2.0 was selected. The game mechanics of LightBot have a one to one relationship with programming concepts but without using a typed language code (Yaroslavski, 2014). The developer of LightBot explains relevance of the game to programming by splitting the concepts into two groups: (1) programming practices and (2) control-flow. Programming Practices group is subdivided into planning, programming, testing and debugging stages to explain the order in which programmers solve the problem using instruction icons without actual coding being involved. The Control Flow group is made of sequencing instruction (conditional statements), procedures (functions), and loops (including recursion) to deal with the step-by-step sequence of program execution. Figure 3 shows the two concept groups used in LightBot.

Figure 3. LightBot 2.0
The game utilizes a fictional scenario, where players control a robot, whose task is to light all blue tiles in a given walking area. This is done through a set of commands representing basic programming concepts such as sequential execution, functions, recursion and conditional flows. The game was next mapped with the US-6774 curriculum (Figure 4). LightBot contains four stages: Basic, Recursion, Conditionals and Experts. The Basic stage includes program design basics: sequential flow of execution, debug and testing the program as well as functions and procedure. The next stage is Recursion which covers the different types of loops in programming. Conditional is the third stage containing different kinds of complex conditions: if-then-else, when, for and where. The last stage Expert is combination of all previous contents. Each of the stages has six levels with gradually increasing complexity.

Figure 4. Mapping programming module to LightBot stages

Experimental design

The LightBot game was played by both student cohorts in two separate settings in a classroom lab environment. Cohort 1 comprised of 20 students, who had not studied the US-6774 programming module, but had done some basic entry level computing courses. The second cohort had 24 students. Cohort 2 students had completed the US-6774 programming module, so they had some knowledge in programming.

Figure 5 illustrates the experimental design of game play used in this study. The students were told to play the LightBot levels in a specific order to achieve gradual understanding of programming constructs. The order of the game and time set for tasks differed between student cohorts. This was because the game tasks were aligned with basic and advanced concepts of the subject modules for the two groups. The game play design for cohort 1 had steps designed to gradually introduce the complexity of the programming constructs. These students were asked to play Basic level first, because activities defined in the game mechanics are used in subsequent stages. After achieving basic level, cohort 1 students were asked to play the first level of Recursion and Conditional. The reason behind this game order was to capture the complete programming curricula defined in the game. There was also a subject-related question in the feedback form to assess their learning after playing the game, since these students had yet not attended programming classes. In Cohort 2, students were explicitly asked to play the game in same order as has been designed by the game manufacturer. This was because these students were already familiar with the programming constructs, so could relate to the game mechanics. However, the Expert stage was optional to play for both student cohorts although students were allowed to play this stage only after completing the game play order.

Feedback was collected in two different ways: immediately after the game play and after completion of the diploma. In the first stage, both groups were asked to fill an online feedback form immediately after the game play. Since the game was conducted in a classroom environment, the whole population sample of 44 responses were obtained in
stage one. Both qualitative and quantitative data were collected from the online survey. Qualitative data included open-ended questions to capture student opinion in their own words. Quantitative data was collected using a 5 point Likert scale questionnaire. Questions were mostly aimed at understanding student perceptions of fun elements in the game, and whether they found the game difficult, or whether concepts on loops, conditionals or recursion were clear after the game had been played. Only students in cohort 1 were asked a technical question pertaining to learning of programming constructs based on the contents of the game.

Cohort 1 - (20 students : Played game before attending programming classes)  
Cohort 2 - (24 students: Played game after attending programming classes)

The second stage of data collection involved a paper based survey, which was given to students after completion of their level 5 diploma in computing. In this paper based survey, one third of the sample population size, which is 15 students, responded. This response rate of more than 30% is comparatively high for a digital survey according to the statistical survey literature (Kaplowitz, Hadlock & Levine, 2004). The survey questionnaire had both quantitative and open-ended qualitative questions to gauge overall student experiences. Questions were aimed toward understanding how students’ perceived game based learning now that they had completed their study. Students were asked: whether or not they had downloaded the game on their personal devices after the classroom activity, if they had discussed any of the gaming elements amongst their peers, and if the gaming constructs had assisted them in any manner for their final exam preparation.
Discussion of the experimental data gathered post game play

Data in stage one was collected immediately after the classroom game activity, so student responses were not blurred with recollections from some past event. Accordingly, survey responses noted the present-day and present-time nuances of students who had just experienced a break-away from the traditional classroom teaching methods. Feedback of all 44 participants was collected through a survey. Responses indicate that overall, both cohorts considered game based learning approach effective and helpful to learn programming concepts. Raw data related to the two student cohort findings are briefly discussed next to give the reader a brief outline of the student feedback.

Cohort 1

Of the 20 students in this cohort, 12 students had prior experience of playing educational games, though none had played LightBot. The game modes related to puzzle and adventure were considered more interesting than other modes (i.e., role play and sports). In terms of programming constructs, sequential logic flow involving functions (or basic stage) was rated as easy, while recursion logic was rated as moderately difficult, and conditional (or advanced stage) as very difficult. For the question related to the fun element in the game, half of the class rated the game as "good fun." With regard to the technical question assessing their learning through game play, 13 of the 20 students could answer the question correctly. Overall, the students perceived programming to be interesting, and ranked the game effective in their understanding of concepts to programming (functions, recursion and conditionals).

In the open-ended question where students were asked to describe their experience in programming, student responses varied from "boring" to "interesting." Positive feedback included comments such as "I was fearful of programming, but now it does not seem so bad," "I enjoyed recursion part – it was so brain storming" and "When I started, it was boring, but once I achieved levels, I wanted to go ahead, and now I understand what recursion is.” Other positive terms such as “amazing,” “interesting” and “fun” were sprinkled across the feedback form. However 20% of the class response was not positive about the experience and educational games in general. The open-ended text answers from these students included responses such as: “I don’t like such games,” “Programming is horrible,” “I would prefer to play this game after I have learnt programming” and “It was too boring.”

Cohort 2

The first question we asked was in regard to the levels they had achieved in different stages. Overall, the whole class had completed higher levels in Basic and Recursion, but had achieved lower levels in Conditionals. Largely, the group agreed that the game was effective in learning programming concepts and playing the game had brought clarity to some of the earlier taught concepts. There was also much agreement on including gaming elements in the curriculum, as these elements were considered relevant to what students had learnt in their previously taught modules. However, a small percentage of the students (17%) found the game to be confusing. In the open-ended question for cohort 2, the positive feedback contained such gems as: “It helped me refresh my programming skills,” “I liked the logic,” “The game was pretty enjoyable,” “Very relevant,” “Makes me more confident” and “Good leisure time.” The negative comments included: “Did not help me,” “bit confusing” and “I don’t know if I liked it or not.”

Over both cohorts the general consensus about the impact of game based learning for understanding of programming constructs was overwhelmingly positive.

Discussion of the experimental data gathered after diploma completion

The purpose of data collection after diploma completion was to capture student reflections on the game based learning strategy after completion of the study. At this stage, some time had passed since the game based learning experiment was conducted, so the students could reflect more dispassionately over the alignment of the game elements with subject content. All students had completed the programming study module, so the data collection in stage two was retrospective. Further, the two cohorts were no longer separate since they had all gone through the programming module and some had even completed the NDC study programme. A few students had left PlayIT after
completion of their diploma course, while others had progressed to the next level of study. The survey at this stage was paper-based; a total of 15 student responses were collected.

Participants were invited to give feedback on whether GBL played any part during the course of their study. Of the 15 students who answered the survey, 6 students had downloaded the LightBot game on their personal devices. All students said they had discussed the game constructs with someone, for instance, with other classmates, family, teachers or friends. Some other students had used excerpts of the LightBot game in their oral presentation assignment. Some comments in the feedback form: “I went over the game again and again, which put my confidence up for assignment presentation” and “I included this game to slides, teacher asked me to explain some words, which I did. Am feeling awesome.” Eleven students said that they understood the basics about programming constructs better after playing the game. Comments in response to a related open-ended question were: “It helped me think logically,” “Ya, it explains the concept of programming,” and “I feel more relaxed about programming.” One student remarked on how the learning activity helped in explaining the gaming elements to others: “I have [an] online account for [this] game. I teach game to [other] members to get points and free passes.”

Analysis of the data

This section gives a detailed analysis of data collected. Raw data collected in the survey post game play was analysed statistically to understand if any relationship exists between nominal or ordinal variables. Nominal variables are based on fixed categorical values like nationality which can be American, Chinese, and Indian etc. This study included only two sets of nominal/dichotomous variables where students were asked if they had played any educational game before our experiment or not and if they found the game confusing or not.

<table>
<thead>
<tr>
<th>Table 1. Cohort 1 – Data analysis using $Y_a$ and $Y_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable 1</td>
</tr>
<tr>
<td>Degree of fun while playing the game</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Degree of agreement for the statement: “This game is helpful/effective to learn programming”</td>
</tr>
<tr>
<td>Degree of overall understanding</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Degree of overall difficulty of game</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Other aspects of GBL experiment were ranked using the Likert scale from largest to smallest: degree of fun, degree of difficulty, and degree of overall understanding. The data thus collected from Likert scale is ordinal, as it refers to ranked data. Spearman’s correlation coefficient method ($\rho_s$) and rank biserial correlation coefficient ($\rho_{rb}$) method have been employed to analyse any sort of relationship between the two diverse data sets for the two student cohorts in stage one. Relations are first described as positive or negative, and then can be ranked in the following order: perfect ($\rho = 1$), strong ($\rho >= 0.7$), substantial ($\rho > 0.5$), moderate ($\rho >= 0.3$), weak ($\rho >= 0.1$) or none ($\rho > 0.0$) (Jackson 2011; Miller 1998). Statistical analysis of the feedback data indicates magnitude of relationship for the following variables as shown in the Table 1 (for cohort 1) and Table 2 (for cohort 2).

Table 1 shows statistical figures for cohort 1 students. Spearman’s correlation coefficient method ($\rho_s$) is used for ordinal values in which students ranked their perceptions of the GBL experience using Likert scale, and rank biserial correlation coefficient ($\rho_{rb}$) is used for nominal values in which students were asked whether they had previously played any educational games to which they could answer either “yes” or “no.”

Table 2 shows statistical figures for cohort 2 students. Spearman’s correlation coefficient method ($\rho_s$) is used for ordinal values in which students ranked their perceptions of the GBL experience using Likert scale, and rank biserial correlation coefficient ($\rho_{rb}$) is used for nominal values in which students were asked whether they found the game confusing or not to which they could answer either “yes” or “no.” The implications of these statistical data values are further discussed in the next section.
After completion of the programming study module at PlayIT within the NDC course structure, the overall student performance was evaluated. To put this in perspective, the examination process for each module within the NDC study programme is explained. Students are allowed at most three attempts to pass any module. If a student is not successful in the first attempt, then the student can re-sit the exam in a second and third attempt. However, if all three attempts are unsuccessful, then the student is not allowed a fourth attempt and is considered failed. Table 3 gives an overview of the how the students from the experiment fared in the final exam held in August 2014.

<table>
<thead>
<tr>
<th>Passed in 1st attempt</th>
<th>All Students</th>
<th>Non-Participants</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passed in 2nd attempt</td>
<td>11</td>
<td>6</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Passed in 3rd attempt</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Failed</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>27</td>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>

A total of 71 students were examined in the US-6774 programming module. Of these, 65 students successfully completed the programming module. Further, from the two cohorts who participated in this study, 43 students successfully completed their study in the programming module (US-6774), of which 38 students had passed in the first attempt. Table 3 data are further discussed in the next section.

Discussion

This study has used game based learning approach alongside current teaching methods to engage students and bring about active learning for one subject module in an introductory ICT course. The study identified an educational game (Light Bot 2.0) which covered the core subject areas of a level 5 programming module. This was to encourage student interest by making the learning experience fun. In the game, an animated robot utilizes logical skills associated with programming constructs to light up blue tiles as the robot moves from one place to another. When the player applies correct programming logic rules, the robot can move on tiles which light up with each correct move. In this way, the game gives instant feedback to the player for each correct move, which in turn motivates the player, who tries to light more tiles while the game slowly increases in complexity. Two student cohorts were selected in stage one of this study. The first cohort had limited knowledge of programming and had not completed the level 5 programming module. The second cohort had recently completed this programming module and was at a later stage of study in this same level 5 curriculum.

The findings indicate that students from both cohorts enjoyed playing the game and indicated that games are effective in learning of some programming constructs (e.g., functions, procedures, conditionals and recursions). The feedback from cohort 1 shows that after playing the game, students perceived programming to be interesting. The game created a positive attitude towards studying programming for students who had yet not started the programming module. The more students found the game enjoyable, the more they considered it an effective way to learn programming ($\gamma = 0.761$). Data showed that students’ perceptions on how interesting they consider programming to be was strongly related to the proportion of their enjoyment in the game ($\gamma = 0.700$). After playing the game, students felt gaming elements to be effective way to learn the programming concepts ($\gamma = 0.665$). Substantially positive moderate relationships were also found between degree of overall understanding to the fun element ($\gamma = 0.525$), and also between degree of overall understanding to perception of “programming would be interesting” ($\gamma = 0.597$). Overall 65% of the students answered the assessment question related to programming constructs correctly. However, the data showed no significant relationships between enjoyment and in the level of difficulty or in the level of understanding. Findings suggest the more students understand a topic the more they think the game is effective and the topics are interesting. This finding supports using games which introduce course topics in an easy manner so that students are motivated to learn further. Answers to the open ended question also support this i.e., one student said that he was fearful of programming, but after playing the game he felt more positive towards programming. However, few students from cohort 1 found the game boring, and said that they would have preferred to play the game after completing the programming module.
The findings from cohort 2 also indicate that students consider educational games a very effective way for applying programming concepts. These students had completed their programming module, and enjoyed applying these skills to a gaming environment. Positive substantial relationships exist between relevance of the gaming elements to programming module ($\gamma_s = 0.543$) and it being included as a learning activity in the curriculum ($\gamma_s = 0.429$). The students in cohort 2 had achieved higher stages during game play than cohort 1, although 17% of students found the game to be rather confusing. Most of the students in this cohort said they enjoyed playing the game more than relating it to programming concepts. However, the game also helped them revise their taught concepts in an enjoyable way. We asked students to voluntarily share their game scores with their peers in the classroom, however only 67% shared their scores.

Most of the students preferred puzzle and adventure type of games, as this stimulated them to think along constrained gaming boundaries. Educational games thus encourage players to apply their logic and reasoning to challenging situations. Students in cohort 2 were more likely to want to try out new thought-provoking moves in stricter game settings, which they may not have tried in a directed teaching and learning environment. Tham and Tham’s (2012) study supports these findings, as they showed dramatic increase in students interest in the course when game based learning was implemented. Overall responses show cohort 1 students’ to be more enthusiastic than cohort 2 students, as we get higher degree of positive correlation for gaming experience with the fun element for cohort 1. This may be because the cohort 1 students were at an earlier phase of study in the course, while the cohort 2 students were nearing completion of the course. The cohort 2 students were busy in preparations for their final assessments for all course modules including programming module at the time of this experiment. The cohort 2 students were also trying to relate the game to programming concepts, rather than simply enjoy the game play with learning as a side product. This could be a topic for further research.

After completion of the diploma, we obtained two datasets. One dataset from the paper based survey, a retrospective view of how students perceived the game based learning approach for their programming module. The second dataset comprised all student results for their final assessment. While only one third of the students participated in the post completion paper survey, the open-ended nature of the questionnaire gives an insightful picture of how those students felt about having gaming elements in classroom teaching. Students rated the game as useful and fun-filled learning strategy. They could visualize programming constructs with animated movements made by the robot on a tiled walking area. Some of the responding students had used examples from the game for the oral presentation assessment component of their final exam. This form of participation where students interact by giving verbal explanations to demonstrate their knowledge, further illustrates enhancement of students’ cognitive learning abilities (Tao, Yeh & Hung, 2015). Students said they felt “awesome” when discussing programming. They related topics with animated movements of the robot in the game, revealing their involvement and engagement during the learning process. The positive respondents to this survey show a high level of emotional engagement.

Research in this area has focussed on understanding student emotions, since emotions are closely related to student learning and represent a key factor affecting student results (Cabada et al., 2012). Results indicate that 86% of study participants passed the programming module in the first attempt compared with 44% of non-participants. The passing results from student final assessments do not show a significant difference between study participants and rest of the class. Other factors related to student’s attributes such as self-study, attendance and interest may have contributed to higher success levels. Nonetheless, the overall findings indicate that GBL is a useful pedagogical approach, which may contribute to learning difficult concepts.

**Conclusion and future scope**

We have applied an innovative way to bring about active learning in classrooms through use of educational games. Suggestions from tutors helped in identifying a subject module considered to be difficult by students. This provided us with an opportunity to apply gaming elements to introductory programming within a classroom environment. Students pursuing a diploma computing course were selected for this study. We applied the game based strategy to one group of students who had no prior knowledge of programming, and to another group who had recently completed the programming module. In this manner, we did not set boundaries to when game based learning should be initiated. Our findings indicate that GBL is a useful learning strategy both before subject is taught and after subject has been taught, but with a slight bias toward after subject has been taught. The GBL experiment showed us that students could be actively engaged in applying programming principles with defined gaming steps. The majority
of participating students agreed that gaming approaches to learning can make classroom environments more fun and also make an effective way to grasp some of the difficult concepts.

This study has demonstrated the effective use of GBL as a teaching and learning activity. Students felt confident about practicing the use of programming constructs in a game scenario and were eager to help others in understanding the game strategy. In applied fields of study such as ICT, the inclusion of gaming elements with traditional teaching practices will bring about more active learning. This will be beneficial for tutors as well as students because games could enable students to grasp technology based applications quickly in a more enjoyable learning environment. This study adds to ongoing teaching and learning pedagogies. This could lead to further research in designing of ICT education curriculum, where learning outcomes of different subject modules could be mapped to related gaming elements, to bring about gradual learning, as games moved from basic (easy) to advanced (complicated) levels.

This study has several limitations. It cannot be said with certainty the effect of GBL with other variables such as the success level, social interactions and self-assessments (Huang & Soman, 2013). Another limitation of this study is that the game (LightBot) covered only an introductory course. Advanced programming would require a more complicated and intensive game. The other limitation is the low response rate in the final paper survey.

The authors believe that traditional classroom teaching cannot be replaced since teachers play both an educator and a mentoring role; the addition of GBL to development of pedagogical activities will enhance the teaching and learning experience. To this end further research could include examination of enthusiasm and emotional engagement in teaching and learning in ICT.

References


KAPEAN: Understanding Affective States of Children with ADHD

Fernando Martínez1*, Claudia Barraza1, Nimrod González2 and Juan González3

1Facultad de Ingeniería, Universidad Autónoma de Chihuahua, México // 2Centro Nacional de Investigación y Desarrollo Tecnológico, México // fmartine@uach.mx // cbarraza@uach.mx // nimrod@cenidet.edu.mx // gabriel@cenidet.edu.mx

*Corresponding author

ABSTRACT

Affective computing seeks to create computational systems that adapt content and resources according to the affective states of the users. However, the detection of the user’s affection such as motivation and emotion is challenging especially when an attention problem is present. An approach to convey learning resources to children with learning problems is by means of game-based tools. The first step towards building these ludic experiences is to identify the skills and limitations of potential users. In this work we describe a case based study carried out to observe and get understanding of the level of attention, cognition and memory of children with ADHD (Attention Deficit Hyperactivity Disorder). Fifteen children were invited to play physical and digital versions of the Stroop and Flanker tests as well as memory-based games. The interaction with KAPEAN is allowed with both a mouse and a hand gestures recognition device. To collect data from the user’s experiences the system integrates a webcam, a Microsoft Kinect device, a video camera, and an Emotiv EEG headset. Data was analyzed offline in order to study which affective states are present while children with ADHD played with KAPEAN.

Keywords

Ludic experiences, Affective states, Attention deficit disorder, Learning difficulties

Introduction

The school’s experience for children with ADHD is challenging. This group of children presents problems to meet cognitive and behavioral levels expected in children of the same age and development (Romero-Ayuso, Maestú, González-Marqués, Romo-Barrientos & Andrade, 2006). Children diagnosed with ADHD present cognitive problems that make scholarly activities challenging. Because these children struggle identifying individual graphemes or constructing meaningful words they find it difficult to complete reading assignments. There are issues with their short time memory that contributes to their dyslexia or dysgraphia diagnosis (Pineda, 2009). Low average marks, failed grades, expulsions, dropout rates, and lower rate of undergraduate completion characterize the academic difficulties faced by children with ADHD (U.S. Department of Education, 2015). In addition, ADHD children present persistent symptoms related to inattention, hyperactivity and impulsivity that make it difficult to be socially integrated. Sometimes they adopt a self-defense position and are not aware of their physical strength that sometimes causes that a peer ends hurt, and thereof ignored or rejected for their schoolmates. For these children, teachers have to implement different approaches when assigning activities aiming to gain a student’s interest and encourage peer interactions. In that regard some efforts have been focused to using digital technologies as the medium to create a better accepted learning experience, in some cases, for instance, digital technology which has been seen as a resource that can help teachers to deliver course content in a social story way.

KAPEAN similarly uses a technology-based setting to let the child play with educational resources. While playing children practices some of their memory, attention and reasoning skills. In addition, KAPEAN collects information about the affective states that are present during the educational game as it aims to get understanding of how to take into account children emotions for future designs of adaptive tutoring tools. In the following sections we first review the literature that indicates the interest for developing affection-based adaptive learning tools. Next we describe the methodology followed to design the case study. Then a discussion of results from playing with KAPEAN is presented. Finally, conclusions and the future work for KAPEAN are given.

Literature review

At the school setting children are expected to achieve tasks with minimal distraction in such a way that pupils can acquire information, complete assignments and participate in classrooms activities and discussions. This level of
behavior is most of the time however unachievable for children with ADHD. These children exhibit behaviors that make it difficult to maintain interactions with teachers and with peers. For this group of children the American Psychiatric Association defines a combination of behaviors that can be grouped into poor sustained attention and hyperactivity-impulsiveness (Jackson, 2004). Furthermore, children with ADHD can present other behavioral problems or learning disabilities. The design of successful instructional strategies and practices require from the teacher a careful study of the child’s needs such as how, when and why the child is inattentive, impulsive and hyperactive. In that regard we observed that some work proposed the integration of digital technologies such as computers and mobile devices as the medium to deliver instructional resources. The use of multimedia applications can offer different advantages from its physical resource counterpart, especially when considering that children are within a digital landscape. Many children today stay at home watching TV, listening to music, playing videogames, chatting or surfing the Web. Xu, Reid and Steckelberg (2002) offers a review of the extent to which technology could be seen as effective tools to support computer-assisted instruction, cognitive training, biofeedback training, assessment, and behavior modification, something children with ADHD can positively be affected with. The use of computers in academic contexts is linked to increasing motivation, time spent on task, attention and skills reinforcement. Digital stories, for instance, have demonstrated to be effective on teaching social skills (More, 2008). If digital instruction is also fun then we can create a motivating learning environment. Besides having fun and learning while playing students could try different roles, experimenting without breaking things, taking challenges and achieving better scores, and reflect about certain conflict situations. Teachers on the other hand can introduce a new learning technique, address complex learning subjects, and create new ways for communication and social integration (Pivec, 2007).

Perhaps the most ambitious effort towards designing digital instructional resources is the creation of adaptive educational tools, those that can take into account children’s affection states. For instance, Cabada, Estrada, Hernández and Bustillos (2014) proposed a system for personalizing the learning of mathematics. Information from emotional states and from previous student’s achievements is processed by an intelligent tutoring system, which will regulate the complexity of the following educational assignments. In the ALING project Peirce, Conlan and Wade (2008) show how gaming and educational targets are combined to provide personalized supportive learning experiences. The combination of the approaches proposed by previous work would yield adaptive educational gaming, experiences that must enable learners to take challenges, encourage their curiosity, nurturing their fantasy, and motivate their confidence to control the flow of the learning process. Although there is a great interest in the design of instructional tools that control the delivery of content based on emotional and cognitive states of the user, there are still open questions about the level of detection of possible human emotions present when using technology. One critical factor of this kind of resource is the provision of a learner’s motivation by detecting the person’s emotions and other mental states. Kotsia, Zafeiriou and Fotopoulos (2013) reflect on that by reviewing the available technologies that can be used to collect motor (Karg et al., 2013), physiological (Sanghvi et al., 2011), gestural (Pantic, 2014) or neuronal (Liu, Sourina & Nguyen, 2010) data and which would be processed to get some hints of the user’s affective states. In addition, it is possible to observe work arguing on what criterion can define real affection-based adaptive learning resources. For instance, Calvo and D’Mello (2010) refers to the importance of distinguishing and differentiating the type of affective computing application of interest such as: systems that detect emotions from the user, systems that express what a human would perceive as an emotion, and systems that actually feel an emotion. In that regard, KAPEAN has been instrumented to understand and study the emotions expressed by children with ADHD from their interaction with a game that help them practice their cognitive, attention and reasoning skills. According to the work reported by Roseló et al. (2003) school life is challenging for these children if considered that 88.8% presents stress, 44.4% observe frustration, and 75% shows unsocial behavior.

**Methodology**

As illustrated in previous sections children with ADHD find it difficult to complete assignments at the school and as a result they fail to get good evaluations. Furthermore, a child with such a profile struggles developing social relationships with his or her peers, or sometimes excluded from the schoolmates groups. As a result, the child’s self-esteem is affected. In this section we start by describing the methodology design for the case study aiming to understand the cognition and affective states of children with ADHD. Qualitative and quantitative techniques are instrumented to collect participants’ attachments to KAPEAN. We seek to answer two hypotheses: To what extent it is possible to observe and get understanding of mental functions such as perception, attention, memory, emotions and...
reasoning, and to what extent could technology be the medium to capture the attention of this group of children within educational contexts?

**Conducting the case study**

First, for a period of four months we visit the “Instituto José David,” institution that provides therapy for children with disabilities, to observe behavior and learning challenges of children with ADHD. Then, we reviewed with the therapists the tools and resources they have at hand to offer therapy support, and the common strategies that are used to plan rehabilitation activities. As a result of the meetings with the therapists we come up with the concept of KAPEAN, a tool that could help better understand the level of cognitive, attention and behavioral impairments of this group of children, which could help the IJD offer better and appropriate instructional methodologies and interventions. Next we instrumented a series of game sessions that children attended to play with KAPEAN. Data was collected from the technology layer and from the after games survey, and analyzed to identify how children performs and completes the games, how do they behave while playing, and what affective states were present during their interaction with such a tool. The following sections offer a better detail of the orchestration for the KAPEAN’s gaming experience.

**Sample selection**

The population sample consists of children directly invited to this study and from children that receive therapy for communication and behavior at the “Instituto José David” (IJD), located in Chihuahua’s city in Mexico. This institution is a civil association that makes efforts to help children with either neuropsychiatric (ADD) or neurodevelopmental disorder (autism). Parents and children assisted to a meeting where they were informed about the aim of the project and the children’s profile requirements as well as the protocols that had to be met in order to be part of the test group. The profile requirements, for instance, considered the child’s age and communication skills. The protocol stated that children had to take a profiling test at the IJD, and a neurological test at the “Centro de Atención Universitario” which is located at the Autonomous University of Chihuahua. Once children and parents fulfill the requirements children were accepted to participate in the project.

**The setting**

The “Instituto José David” provided two 3x3 square meters Gesell rooms to run the experiment. The rooms allowed observers from the outside without affecting the development of the experiments. In each room provided by the institution there was a projector, a webcam, a desktop computer, a Microsoft Kinect device and a leap motion device. The computers within the room were used to reproduce the KAPEAN’s mini-games and to collect data from the interaction device (a leap motion or a mouse), as well as from the Kinect device. Two extra computers were installed outside the Gesell room and on these computers data from the EEG headset, and video data from the webcam are captured. The emotive visualization tools and the Camtasia studio were also running on these computers.

**Game objects**

KAPEAN aims to offer technology-based ludic experiences to get understanding of the behavior that children expose when an educational task is assigned. Each of the ludic learning activities integrating KAPEAN, see Figure 1, aims to reinforce attention, cognition, memory, visouspatial skills:

- **Reasoning** - Based on the Stroop effect experiment the game asks children to read aloud the name of a color written a different color ink. The Stroop effect is commonly used in psychological evaluation and requires from the user to reason about the semantic meaning of the word and the color of the word. The Stroop test has been applied to study how humans select appropriate responses to resolve conflicts. With KAPEAN the participant uses his/her rational thinking because it is important to separate the two information channels: the meaning of the word and the color it is painted.
- **Attention** - Based on the Flanker test the game demands from children to identify a target that is flanked by non-target stimuli. The test allows studying the cognition process involved in the recognition of targets in the
presence of distracting information. The Flanker test has been applied to study how the human discerns incongruent and congruent stimuli. With KAPEAN the player must focus his/her efforts to discard objects that add noise to the task of finding the matching cards.

- Memory - The short term memory seeks to understand memory skills from children with ADHD, who show inconsistent performance during this type of tests. Some cards are presented to the player during a few seconds and immediately s/he has to remember these cards from an unordered and larger set of cards.
- Spatial memory -  Visuospatial working memory is the responsible for storing visual and color information. Children are required to summarize and recall information about the spatial location of objects on a grid.

```
Figure 1. The four games that form part of KAPEAN: memory (left), attention (center-top), reasoning (center-bottom) and visuospatial (right)
```

Recording and tracking the user’s interaction

A camera recorded video footage for each one of the game sessions. This video material was used for both assessing the completion of the assignment and studying affection states that can be detected from body gestures or from any phrase the child spoke out. There is also a webcam that recorded face gestures and helps, together with the EEG data, to review the child affective states. Data collected from a Microsoft Kinect device was used to review interaction experiences with the leap motion device, which allows the use of hand gestures to interact with computers. From the pre-test group interaction experiences we observed that children with ADHD were careless with the field of view the leap motion has available to detect gestures. The Kinect thus helps to get a record of the number of times the child’s hand was out of the focus of the leap technology.

Children can play the KAPEAN digital games using, one at a time, the mouse and the leap motion (www.leapmotion.com), device that uses two cameras and three infrared leds to track hands and fingers gestures. The use of the mouse to play games enables us to identify to what extent children are familiar with computers. The use of the leap motion enable us to understand to what extent children can adopt interfaces that allows a more natural interaction with computers, using hand gestures in this case. For the interaction children need to use a pair of fingers and make them behave as tweezers to recreate the mouse’s selection event.

EEG data

In order to complement the video material used to identify affective states we used an EEG Emotiv EPOC + headset. The user wears the headset, which sends EEG data to the central computer. The EPOC headset has 16 electrodes to collect electrical activity along the scalp of a human being. The headset uses the international system 10-20 to collect data from the cranium’s key nodes: nasion (near the frontal lobe), inion (occipital lobe), and the pre-auricular points.
Table 1, shows the information that is registered from brain activity during the interaction of a participant with the KAPEAN’s game.

**Table 1.** Position of the Emotive EEG headset electrodes, and the brain areas being monitored

<table>
<thead>
<tr>
<th>Brain area</th>
<th>Function</th>
<th>Left hemisphere electrodes</th>
<th>Medium line electrodes</th>
<th>Right hemisphere electrodes</th>
<th>KAPEAN games</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontopolar</td>
<td>Executive; cognition.</td>
<td>FP1</td>
<td>FP2</td>
<td></td>
<td>Attention &amp; Reasoning</td>
</tr>
<tr>
<td>Frontal</td>
<td>Executive, reasoning.</td>
<td>F3</td>
<td>Fz</td>
<td>F2</td>
<td>Attention &amp; Reasoning</td>
</tr>
<tr>
<td>Fronto-temporal</td>
<td>Emotions; attention; motor ability.</td>
<td>F7 C3</td>
<td>Cz</td>
<td>F8 C4</td>
<td>Attention</td>
</tr>
<tr>
<td>Medial Temporal and Parietal</td>
<td>Episodic memory; visual memories; emotions.</td>
<td>T3 P3</td>
<td>Pz</td>
<td>T4 P4</td>
<td>Memory &amp; Visuospatial</td>
</tr>
<tr>
<td>Posterior Temporal and Occipital</td>
<td>Orientation; spatial-frequency; visual world.</td>
<td>T5 O1</td>
<td></td>
<td>T6 O2</td>
<td>Visuospatial</td>
</tr>
</tbody>
</table>

The data provided by the EEG headset is processed to obtain the neuro-rhythmic signals that represent the electrical activity at the human being scalp. In order to train the classifier of emotional states, a training session was prepared before the child interacts with KAPEAN. The training session consisted of a pair of experiments. The first focused on the collection of EEG data when the child was asked to develop an attention task, which consisted on solving puzzles using a tangram game. For the second session the child was asked to remain relaxed within the room and when that mental state was identified then a distractor event was inserted. The total of the reproduced distractions were three and could include, for instance, make a mobile phone ringing within the room, or suddenly knocking on the door. The EEG data collected from the training session is fed into a naïve Bayes classifier, which uses the cerebral signals alpha, beta, theta and delta to differentiate meditation and engagement states when the child plays with KAPEAN.

**Playing with KAPEAN**

When the child arrives to the experiment’s room the emotive headset must be collocated on the child’s head making sure that the headset’s electrodes have good contact with the skull. Next a psychologist asks the child for some information about his/her mood and emotional states, which help us to take this information into account when analyzing the video material. After that, the game is setup and applied.

*Figure 2. Interacting with KAPEAN using the leap motion device (The player is wearing the Emotiv headset and some of the affective states are displayed (left))*
There are three kinds of sessions orchestrated to interact with KAPEAN. The child can play using a physical resource or with its digital version in which case the interaction device can be either the mouse or the leap motion. If playing with the physical version of a game then an adult takes the role of a tutoring companion within the room. The child’s skills to complete the game define the time of the duration of it, which cannot last more than 10 minutes. If playing the digital version of a game then the child is left alone within the room and s/he plays on its own. The digital games are programmed to last 10 minutes. Once the game session expires the headset is retired from the child’s head and a satisfaction survey is applied. Synchronized with the beginning of the game, the cameras start video recording and so the emotive tools. These helps with the visualization of the performance of the EEG headset and the behavior of EEG rhythms in real time, see Figure 2. As indicated previously, the Kinect device helps to record information of the moments when the child is not aware of that his/her hand might be out of the field view of the leap motion device, see Figure 3.

![Figure 3. Playing the reasoning game (The webcam, the Microsoft Kinect device and the video camera track the player’s activity)](image)

Identification of affective states

Three qualitative analyses were undertaken in order to identify affective states that are present when the player interacted with KAPEAN. First, the psychologist observed the gesticulation made by participants along the 10 minutes of the game session. Every minute an annotation was made on the observational user experience’s record. From the observation of the participants’ gestures the specialist can annotate whether the participant seemed engaged, tired, desperate, or even annoyed. On this sheet there are other established categories with three-to-four multiple options each that the specialist can tick to indicate the mood state, the body and hand posture, or any other gesticulation.

![Figure 4. The usability test observation code, UTOC (a) and the Emotiv affective suite (b) are two of the resources that supported the evaluation of the participants’ affective states](image)
The second analysis considers the offline review of video material recorded with the video camera. The usability test observation code (UTOC) for this part of the qualitative analysis offers more detailed categories to evaluate the user’s experience with KAPEAN. Figure 4a, shows an extract of the UTOC template. The third analysis consists of the affective states’ identification using the emotive affective suite, Figure 4b. This tool has a user interface displaying the engagement, frustration, meditation and excitement affective states levels in a scale from 0 to 1. Using the Camtasia suite®, we review minute by minute the predominant affective state and compare the results with these obtained from the UTOC and the one generated by the psychologist. That is, the outcomes from the three observational analyses are evaluated together and the most preponderant affective state on each minute is taken as the dominant affective state.

**Meditation and disengagement**

In addition to the identification of affective states it was also important to get understanding of the level of engagement children with ADHD can held when developing an educational activity. In order to do that, we use data collected with the EEG headset and fed a classifier which outcome indicates the levels of the EEG rhythms, alpha, beta, theta and delta, and the correspondent levels of meditation and disengagement, Figure 5.

![Classification](image)

*Figure 5.* The output from the naïve Bayes classifier indicating moments of meditation and disengagement (Every 20 samples EEG rhythms are processed)

**Results**

In this section we present some results from the children’s experiences with KAPEAN. As indicated earlier in this work our interest is to understand some of the affective states that children with ADHD expose when working with an assignment. Based on what literature reports regarding attention and short memory limitations that this group of children often presents we built a series of digital games each one targeting a particular learning issue. The memory game tested the child’s skills to remember objects, the visuospatial game assesses how the child processed visual information to work out spatial relationships among objects, the attention game demands the application of logic skills, and the reasoning game requires from the child the application of rational thinking. For the sake of space we use the performance of a couple of participants with the visuospatial game to discuss about affective states. The experience of some children with the memory game to illustrate cognitive demands that might negatively contribute to the affective states, and some of the children’s performance with the attention game to identify levels of the engagement affective state.
Playing with the visuospatial game using the hand gestures recognition device

Remember that the leap motion can track gestures done with the hands. The ability of this device to detect hands events is limited to 20 cm maximum on any axis extending from the center of the device. Thus the player must be aware of the detection area in order to not lose the “mouse” presence on the displaying area. Physical and cognitive efforts, such as the one already mentioned, are drawn from the qualitative analysis of affective states, see Figure 6. For the participant N002, for instance, we identified a low level of excitement when s/he played the visuospatial game using the leap motion device. By the end of this game, however, there is a moment of instantaneous excitement that may be associated to the enthusiasm of stopping using this interaction device. The frustration level is above a medium level most of the time with peaks around minute 6 and 9. At minute 6 the player was annoyed due to the difficulties faced with the use of the finger gestures to interact with the game. At minute 9 the player gave up and stops interacting with the game. After reviewing the video footage it was realized that it was hard for this participant to use the leap motion technology. Some of the verbalizations this child expressed during the game were:

“Do not get stuck… listen to me… where are you… come on come down… it does not like to function… this thing does not let me play… no, no, I can’t it is too much difficult…”

When reviewing the data collected by the leap motion device it indicates that the user is not aware of the sensible area that the leap motion device allows for the hand gestures recognition. File logs recorded with the Kinect device indicated that the child was not aware that it was the hand’s wrist, and not the hand itself, the one that was most of the time on the angle view of the leap motion device. This situation, therefore, create moments of frustration for the child. In addition, it was possible to identify that the participants with highest level of ADHD were the most frustrated with the use of the leap motion. It seems there is some consistency with what literature reports in respect to the fine motor problems present with children with ADHD. Associated with the frustration state, this child seemed to use an additional effort to get throughout the game. The meditation level, light grey line, is most of the time above the other affective states and by half the game it starts crossing the frustration line. By reviewing video material we observed that on the second half of the game the child appears to be guessing instead of reflecting on the right answers.

![Figure 6. Affective states reported by the participant N002 when playing the visuospatial game with the leap motion](image)

![Figure 7. Affective states reported by the participant N008 when playing the visuospatial game with the leap motion](image)
Different from the N002 participant the behavior of the participant N008 seems to be more relaxed during the game session. Figure 7 shows that there are few moments of excitement along the game session, that frustration is only present around minute 3, and that apparently the completion of the game was straightforward. Although it is not clear the level of the player’s enthusiasm, the affective states could reflect that the use of the leap motion seemed to represent not much trouble for this child.

Playing with the memory game using both the mouse and the leap motion device

The use of unfamiliar technology affects the behavior and the attitude of a child to complete an assignment. It has been shown that the use of emergent interaction technologies such as the leap motion device can be challenging for children. Overall it took more time to generate a click using the leap motion than using the mouse. Figure 8 presents the click times used by participants when playing the memory game. For the participant N002, for instance, it took more than a minute to “click” on targets using the leap motion. The video analysis showed that when a child played using the mouse s/he was unaware of the device presence and focusing on clicking the answers. On the other hand, when the child used the leap motion to interact with KAPEAN s/he was completely aware of the device and from time to time got distracted from the task. This was more evident for children that have a profile with fine motor problems.

![Figure 8. Time span required by participants to click on targets when playing the memory game](image)

Playing with the attention game using the mouse

The attention game asked the child to select a target in the presence of distracting information. EEG data collected when children played this is analyzed in order to observe the level of meditation and engagement (or better said disengagement) states. A naïve Bayes classifier reported the moments when the child was focused on the tasks and the moments when the child got distracted. The Figure 9 shows this affective state for the seven children that played the attention game, a high state indicating that the child was focused on the task whereas a low state indicating moments when the child is disengaged from the activity. The participant N017 was the player that was never engaged with the game. After reviewing the psychologist notes we realized that the child arrived to the experiment in bad mood because s/he missed his karate class. Moreover, this child uses medication and on that day it was not administered. Although the child accepted to play, the video material reported that, the child was just watching the wall clock and spinning on the chair. Therefore, the interaction of this participant with the attention game was minimum, perhaps just when the child realized that the end of the game was near. On the contrary the child N009 was all of the time focused on the game.
Conclusion

This work presented the design of a case study built to observe and understand how children with Attention Deficit Disorder and Hyperactivity behave when developing learning tasks. Fifteen children were invited to play games that are often used by therapists to reinforce memory, attention, visuospatial and reasoning skills. Various kinds of technologies helped to record the children interaction with KAPEAN in order to gather information to support the analysis of the users’ mental states. Affection states related to engagement, frustration, meditation and excitement were coded from EEG data, video analysis and scores obtained from the gameplay. Without surprises it was observed that children with ADHD struggle to keep focused with a task. The frustration state appears to be triggered by the lack of experience of the child with the leap motion technology, and also by the games that required the most of their attention. The meditation and the engagement states were used to assess the child behavior during the development of a task. Nevertheless children liked playing digital games than their physical counterpart. Although KAPEAN enables us to understand some of the affective states that a child with ADHD exposes while completing an assignment we would like to improve KAPEAN in order to take into account other mental states such as cognition. We are also working on the automation of detecting emotional states from face gestures. Furthermore, the next version of KAPEAN would seek to adapt itself the educational content delivered to a child based on these mental states. That is, once the system can identify from the EEG’s rhythms that the child is moving to a state of disengagement then the displayed content of the current assignment should be substituted with a content that would keep a sustained attention of the child.

Acknowledgments

We would like to thank the PROMEP for funding the “Platform for the Experimentation with Mobile Technologies” (PEMT) project, at the Autonomous University of Chihuahua. We are also grateful with Magda Maldonado and
Rafael Holguín for their comments and Adrian Muñoz for the help with the coordination of the experiment’s sessions. We appreciate the support given by the Instituto José David and all of the children that kindly participated with this project.

References


Roselló, B., García Castellar, R., Tárraga Mínguez, R., & Mulas, F. (2003). El papel de los padres en el desarrollo y aprendizaje de los niños con trastorno por déficit de atención con hiperactividad [The Role of parents in the development and learning of children with attention deficit hyperactivity disorder]. Revista de Neurología, 36(1), 79-84.


Affective Behavior and Nonverbal Interaction in Collaborative Virtual Environments

Adriana Peña¹‡, Nora Rangel², Mirna Muñoz³, Jezreel Mejia³ and Graciela Lara¹

¹CUCEI, Universidad de Guadalajara, México // ²Centro de Estudios e Investigaciones en Comportamiento, Universidad de Guadalajara, México // ³Centro de Investigación en Matemáticas, Unidad Zacatecas, México // adriana.pena@cucei.udg.mx // norarangel@cucba.udg.mx // mirna.munoz@cimat.mx // jmejia@cimat.mx // graciela.lara@red.cucei.udg.mx

*Corresponding author

ABSTRACT
While a person’s internal state might not be easily inferred through an automatic computer system, within a group, people express themselves through their interaction with others. The group members’ interaction can be then helpful to understand, to certain extent, its members’ affective behavior in any case toward the task at hand. In this context, a Collaborative Virtual Environment for learning represents a resourceful technology where students can interact with the environment and with each other by their graphical representation, i.e., their avatar; and at all once, each participant’s action can be automatically followed. We propose the analysis of nonverbal interaction to identify affective behavior during the accomplishment of collaborative tasks, which can be extended to verbal content. An exploratory study was conducted to understand the possibilities of this approach; the results showed behavior patterns that can be associated with task-focused affective states.

Keywords
Affective learning, Collaborative Virtual Environments, Nonverbal interaction, Collaborative Learning

Introduction

Emotions influence our behavior (Kantor, 1921a; Kantor, 1921b). And as such, during the last two decades, the interest in the affective component in computers has grown under the research area of affective computer; aimed to the study of the relation that involves emotions and computers (Picard, 1997, p. 50). For learning, a special effort has been made to automatize the important tutor capability of the humans, to comprise the emotions of their students in the teaching process (Heyward, 2010; Moridis & Economides, 2008), that is, affective learning. However, due to their measure complexity, the automated sensing of emotions remains as an open issue.

According to Piccard & Daily (2005), the methods to evaluate affect for computer-users include the classic questionnaires. This self-report on emotions presents inconveniences such as interruption, and the fact that the emotional state of a person can change from moment to moment. The evaluation of affective behavior also includes body measures. Both, based on physiological signals through sensors such as for body-worn, EEG (electroencephalogram) or ECG (electrocardiogram) (e.g., Agrafioti, Hatzinakos, & Anderson, 2012; Bamidis, Papadeli, Kourtidou-Papadeli, Pappas, & Vivas, 2004), or through body activity such as facial expressions, posture, hand tension, gestures or vocal expressions (e.g., Ammar, Neji, Alimi, & Gouardèresc, 2010; see Picard & Daily, 2005). Typically, this type of evaluation requires special equipment, and for the user to be greatly aware of it. A third approach is the one based on task-measures, that as indicated by Picard & Daily (2005), is an indirect evaluation that considers that our affective state influences our behavior on subsequent tasks (e.g., Lerner, Small, & Loewenstein, 2004; Liu, Lieberman, & Selker, 2003). Although, this type of measure is not intrusive, because task-measures are indirect, their results are usually applicable in populations and not in individuals (Picard & Daily, 2005).

On the other hand, what a person does is not an expression of internal or innate entities, but a direct effect of what is happening in the environment (Kantor, 1921a; Kantor, 1921b). Almost a hundred years ago, Kantor (1921a; 1921b) was already concerned about the recurrent causal connection between mental and physiological states in the study of emotions. On this regard, Boehner, DePaula, Dourish & Sengers (2007) recommended for affect computing to incorporate an interactionist approach, emotions as constructed through interaction and expression, and not as an objective natural fact. Boehner et al. (2007) proposal is mainly intended for the human-computer interactions (HCI), and not particularly for computer-mediated human interactions.

For Kantor (1929), the elements to consider in the analysis of affective behavior are: the stimulus or events that the individuals face, the individual’s behavioral repertoire, the reaction speed, the physiological conditions of the person,
the familiarity with the stimulus object, the context and the interaction circumstances, and the presence of certain persons in the situation to be analyzed. These last elements turn out to be particularly important when the aim is to analyze the affective component of the individual’s behavior, within a group, while they solve a collaborative task.

People express themselves through their interaction with the others; and in computer-mediated interactions, every user intervention can be recorded and analyzed in time. However, automatic understanding of unstructured dialogue has not being completely accomplished yet (Jermann, Soller, & Lesgold, 2004) or it represents a high computer cost. Nevertheless, in Collaborative Virtual Environments (CVEs) users interact via their graphical representation, their avatar, which includes the capability of the display of nonverbal interaction. Our nonverbal behavior comprises most of what we do but the word meaning, including patterns of verbal interchange like gaps or pauses (Heldner & Edlund, 2010), the objects when they are part of the task at hand or proxemics behavior, to name some among other nonverbal cues.

Nonverbal behavior has a number of studies in the real world (see Knapp & Hall, 2010, pp.18-21) and in the creation of artificial behavior in robots and animation (e.g., Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005; Guye-Vuillème, Capin, Pandzic, Thalmann, & Thalmann, 1998); however, there are few studies of the nonverbal cues people display in CVEs through their avatars. In 1998, Guye-Vuillème, Capin, Pandzic, Thalmann, & Thalmann recognized the importance of the conversion of non-verbal communication to an equivalent in virtual worlds. By presenting a study of the influence of complex embodiments, they emphasized the avatars functions in collaborative situations as: perception, localization, and identification, among others. Particularly for object-focused interactions in virtual environments, Hindmarsh et al. (2002) studied how an immersive desktop CVE system provided participants with the ability to refer and discuss features as well as their interaction around objects. On this regard, according to Schroeder (2011, p. 105), for instrumental tasks in CVEs is important to facilitate joint orientation. Schroeder (2011, p. 105) also argues that in such cases the users seem to be able to ignore the absence of many nonverbal cues using those available.

The observation of nonverbal behavior to identify affective behavior represents a task-measure approach based on interaction. However, by using the own participants’ nonverbal cues as a parameter for measures, these measures are individualized. Though, it is important to point out that the features of CVEs, which are predominantly visual, are proper for collaborative tasks in which people need to center their attention on the space and the objects in it, spatial tasks where co-presence is desirable, otherwise this technology might not be required (Spante, Heldal, Steed, Axelsson, & Schroeder, 2003). Also, unlike in real life, the focus in a CVE will be narrowed on a few things and constantly engaged because there is an ongoing reason for being in it (Schroeder, 2011, p. 45).

Nonverbal interaction cues and affective behavior in learning

As Ben Ammar et al. (2010) pointed out, affect is inextricable bound to learning. Expert teachers are used to recognize and address the emotional state of the learners and based upon that observation, to take actions to impact their learning; placing this skill in computers should result in better computer assistance for the learner (Kort & Reilly, 2002).

For the analysis of the term “emotion” have been included a number of phenomena: emotions, passion, sentiments, affective conduct, motivations, and moods, among others (Rodriguez, 2008). Kantor (1921a; 1921b; 1929) defined emotions as a “no answer” moment, like an action interruption for an overwhelming stimulus; by the time the individual responds, the emotional conduct might be gone. In contrast, sentiments are organized reactions produced by the experience; they functionally correspond with the stimulus or events that the individuals face (Kantor, 1921a; Kantor, 1921b; Kantor, 1929; Ryle, 1949). While sentiments are the affective component of the effective behavior, emotions interfere with that behavior; and both facilitate certain activities, not as a cause but as a disposition (Rodriguez, 2008). Sentiments produce changes in the person and they can create attitude changes toward the stimulus, increasing or decreasing the general function of the person, retarding or accelerating his/her activity and generating less or more interest in something in particular (Kantor, 1921a; Kantor, 1921b; Kantor, 1929).

Learning is not the exception; emotions are related to learning affecting the students’ performance (Kort & Reilly, 2002). Kort and Reilly (2002) described the dynamics of emotional states in the areas of science, math, engineering
and technology, naturally involving failure and a host of associated affective responses. The Kort and Reilly (2002) model is composed of four quadrants associated to affective states as follows:

- Quadrant I: “awe, satisfaction, and curiosity”;
- Quadrant II: “disappointment, puzzlement, and confusion”;
- Quadrant III: “frustration and discard of misconceptions”;
- Quadrant IV: “hopefulness and fresh research”.

According to Kort and Reilly (2002), the students ideally begin in quadrant I or II, because they might be curious or fascinated about a new topic of interest (Quadrant I) or they might be puzzled and motivated to reduce confusion (Quadrant II). Then, as the students construct or test knowledge, if they fail, they will be in Quadrant III where changes are required and understanding what might need to be done or what does not; while making progress, they may move to Quadrant IV with fresh ideas and then they may move to any of the other quadrants, until they solve the problem or complete the task. And, the students could be simultaneously in multiple quadrants, for example, they might be frustrated but also curious about how to proceed. It is worth to mention that it is out of the scope of the paper to theorize about the axes in the quadrants of the Kort and Reilly’s model, that is, their labels of negative to positive affect and the unlearning to constructive learning. Nevertheless, this model manages states that can be recognized by changes in the group nonverbal interaction; and it fits to the accomplishment of an open-ended task, that is, a task with different solution possibilities causing, in some cases, that the participants fail before they find a right solution.

In such a way that the flow (Csikszentmihályi, 1991) during collaboration can be interrupted when the participants find difficulties on taking care of the task or when they disagree on the course of action. This flow interruption can be associated with frustration, puzzlement or confusion versus satisfaction or the determination to finish the task affective behaviors, as in the Kort and Reilly (2002) model.

Now then, we assume that in an engaging situation, due to the emotional state, the implementation of the task can be momentarily interrupted when one of the participants feels that something might be wrong and therefore feels frustrated, puzzled or confused. As a spatial task, the interruption in the implementation could be due to a change in the location of the participant, related to for example, an overview of the scenario; therefore a navigation interruption has to be considered. Afterwards, longer than usual utterances, could represent the moment when the participant explains or discusses his/her point of view on this regard, a reaction to the emotional moment. To our knowledge this alternative has not been explored elsewhere and therefore we decided to go at front with an exploratory study to understand its implications.

### Exploratory study

As aforementioned, we propose the use of certain nonverbal behavior cues as an alternative to measure affective behavior such as frustration, confusion or puzzlement during the accomplishment of an engaging collaborative task. In this first approximation, for simplicity, as the minimum expression of a group dyads were observed. The nonverbal cues selected were the duration of utterances, object manipulation inactivity linked with navigation inactivity. These atypical behaviors of the participant are assumed as an affective moment; then this data is contrasted with what the participants verbally express to verify, at least to a certain point, if they somehow actually correspond to an affective behavior. 

### Participants

Eight classmates undergraduate students, seven male and one female, with ages in the range of 18 to 23 years old, voluntarily participated in the study in exchange of desktop objects they could choose (i.e., pencils, text markers or pens).

### Materials, devices and experimental situation

The sessions took place in an illuminated room, free of distractions. In each of two desks in opposite directions was placed a Dell™ computer model Allienware X51, for oral communication microphones with earphones and the
TeamSpeak™ application was used. Both computers were Internet connected and the participants’ manipulation of objects, navigation in the environment, and starting or ending utterances were automatically registered in a text file.

With the OpenSim™ software and the CtrlAltStudio™ viewer applications, a tridimensional (3D) CVE was adapted where the participants worked in dyads. Both participants could see and interact in the same virtual scenario, but from the point of view that matched their place in the virtual world at any given moment. The sessions were videotaped with the Fraps™ application, it saves the screen as the user sees it, and the screens of both participants were videotaped.

**Design and procedure**

The implemented design allowed comparisons intra and among participants and groups. The participants were divided in four dyads that were randomly assigned. The dyads were exposed to an instructional session for the software application, a demonstration session. Afterwards, each dyad was exposed to a session to solve the experimental task, which consisted in the assembly of several pieces to form a geometric figure, like the one in the Figure 1. During the sessions, the participants could saw that same model formed by plastic pieces.

![Figure 1](image_url)

*Figure 1. The figure to be assembled in plastic pieces*

The scenario was an island in which a number of different color pieces were placed around a rectangular plane as shown in Figure 2. In the CVE each user was graphically represented by a humanoid avatar in first person, that is to say, they could not see themselves in the virtual world although they could see his/her partner. The avatar corresponded to the participant’s gender and had written the participant’s name above the avatar to facilitate identification and communication.

The human-computer interaction was through keys combinations and the mouse. The participants could navigate by earth, for example by walking or running, and by air flying; and select, move or rotate each piece or object. Navigation is not always necessary for object manipulation; objects can be manipulated from the distance if they are at sight. During the session a piece of paper with the keys combination was available in the desks. There was no limited time to solve the task.

During the demonstration the participants were allowed to familiarize with the application for around 3 to 5 minutes before they started the session. Then the next instructions were given to them: “To communicate with each other use the microphone with the earphones. The two of you are going to work together to assemble a figure like this (the figure in plastic was then shown to them and it was also verbally explained), a cross with two levels, each bar of the cross has two cubes. Not all the pieces are necessary to assemble the figure; please assemble it within the white plane. We are here in the next room, give us a call when you finish.”
Data

The software application saves the “X,” “Y” and “Z” coordinates of the position of the objects and the avatars, when the pieces or the avatars are moved or rotated, and when the microphone is deactivated or activated by the voice of the user. Data was treated to calculate: the duration of the utterances, and the inactivity periods for navigation and the accomplishment of the task, that is, the elapsed time when each participant released a piece and he/she grabbed another one.

The outstanding points, those with the longest duration, were calculated for each session for the three indicators: utterance duration, navigation and manipulation of objects inactivity. The calculation was made with the media and the standard deviation of these indicators where the outstanding points or atypical data are the indicators with 5% of probabilities to occur in the upper limit one tailed. As an example, in Figure 3 is shown, the scatter-chart of the
utterance duration of one of the participants of the Dyad 1. The points connect the initiation of an utterance (X-axis) and its duration (Y-axis). The longest utterances are the points on the top of the scatter-chart, marked within a black rectangle.

Data was also treated to see when more than one piece was taken out of the workspace, this represents when the figure was partially disassembled.

The verbally expressed difficulties found during the accomplishment of the task were obtained and classified as follows:

- **Doubts about the figure shape**, when at least one of the participants verbally expressed not to be sure about the final shape of the figure.
- **Doubts about the pieces**, when at least one of the participants verbally expressed his/her concern about not having enough pieces or the proper ones to finish the figure. Sometimes this difficulty was related to doubts about the figure shape.
- **Doubts about the course of action**, when at least one of the participants verbally expressed his/her concern about what his/her partner was doing. Here were observed expressions like “What are you doing?” or “Why are you doing this or that?” or when one of them expressed not being sure if what they were doing was correct.
- **Problems handling the pieces**, when at least one of the participants verbally expressed that he/she could not, for example, lift or rotate a piece. Which might be more related to frustration than to puzzlement or confusion.

Other difficulties founded were:

- **Not dialoging yet**, when the atypical behavior took place before they started to talk to each other. Dialogue is fundamental for collaboration, although people might collaborate in a spatial task without talking to each other that represents a very uncommon situation.
- **They asked for help**, in this particular case the participants were not sure about the final shape of the figure and they decided to ask for help instead of reaching an agreement.
- **Disassemble of part of the figure**, the time when they disassemble part of the figure. No dyad disassembled the whole figure.

**Results**

All dyads completed the task, Dyad 1, 3 and 4 successfully; Dyad 2 did the figure using three, instead of two cubes, for each bar of the cross. The time each session took is shown in minutes and seconds in the second column of the Table 1. All the dyads started the session, once both participants were connected, by moving pieces. The dialogue started later in all cases, as shown in the third column in minutes and seconds.

<table>
<thead>
<tr>
<th>Dyad</th>
<th>Session duration</th>
<th>Dialogue started at</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13:53</td>
<td>01:45</td>
</tr>
<tr>
<td>2</td>
<td>17:19</td>
<td>05:18</td>
</tr>
<tr>
<td>3</td>
<td>15:18</td>
<td>01:55</td>
</tr>
<tr>
<td>4</td>
<td>08:17</td>
<td>00:07</td>
</tr>
</tbody>
</table>

The time of all the atypical behaviors and the verbally expressed difficulties are shown as follows: in Table 2 for Dyad 1, Table 3 for Dyad 2, Table 4 for Dyad 3, and Table 5 for Dyad 4. In the first column of these Tables, (Elapsed Time) is the elapsed time during the session when the atypical behavior took place in minutes and seconds. In the second column (Participant) is a given number to identify each participant (1 or 2) in a dyad. In the third column (Duration) is the duration of the atypical behavior in minutes, seconds and tenths of second. In the fourth column (Atypical behavior) is the type of atypical data: OM for inactivity in manipulation of objects, NA for inactivity in navigation and UT for utterance. When the inactivity in object manipulation corresponds to inactivity also in navigation, or inactivity in navigation corresponds with inactivity in object manipulation, the atypical behavior is marked on its right side with an asterisk “*”. In the fifth column (Related difficulty) is the verbally externalized difficulty that took place in the environment around that time, according to the audio of the session. When the related difficulty was “Problems handling the pieces,” the participant expressing this problem is in parenthesis. When there
is no atypical behavior only this column has data (this happened only once, see Table 4 at 09:54). When there was no verbally expressed difficulty it is denoted with a short slash "-". A row with the time was added to indicate when the dyad disassembled part of the figure. By calculating this data, an automatic analysis can be performed during the collaborative session.

Table 2. Atypical behavior of Dyad 1

<table>
<thead>
<tr>
<th>Elapsed time</th>
<th>Participant</th>
<th>Duration</th>
<th>Atypical behavior</th>
<th>Related difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:45</td>
<td>1</td>
<td>00:04:70</td>
<td>UT</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>01:47</td>
<td>2</td>
<td>05:11:85</td>
<td>NA</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>02:01</td>
<td>1</td>
<td>00:13:93</td>
<td>OM*</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>03:12</td>
<td>1</td>
<td>00:16:12</td>
<td>OM*</td>
<td>Doubts about the pieces</td>
</tr>
<tr>
<td>03:16</td>
<td>1</td>
<td>00:02:31</td>
<td>UT</td>
<td>Doubts about the pieces</td>
</tr>
<tr>
<td>03:17</td>
<td>1</td>
<td>03:56:705</td>
<td>NA</td>
<td>Doubts about the pieces</td>
</tr>
<tr>
<td>03:27</td>
<td>1</td>
<td>00:02:29</td>
<td>UT</td>
<td>Doubts about the pieces</td>
</tr>
<tr>
<td>03:40</td>
<td>1</td>
<td>00:02:57</td>
<td>UT</td>
<td>Doubts about the pieces</td>
</tr>
<tr>
<td>03:55</td>
<td>2</td>
<td>00:21:49</td>
<td>OM*</td>
<td>Doubts about the pieces</td>
</tr>
<tr>
<td>04:05</td>
<td>1</td>
<td>00:28:90</td>
<td>OM*</td>
<td>Doubts about the pieces</td>
</tr>
<tr>
<td>04:53</td>
<td>2</td>
<td>00:02:68</td>
<td>UT</td>
<td>-</td>
</tr>
<tr>
<td>05:14</td>
<td>1</td>
<td>00:02:68</td>
<td>UT</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>05:19</td>
<td>2</td>
<td>00:45:50</td>
<td>OM</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>05:28</td>
<td>1</td>
<td>00:23:53</td>
<td>OM*</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>06:28</td>
<td>1</td>
<td>00:02:37</td>
<td>UT</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>06:36</td>
<td>1</td>
<td>00:02:58</td>
<td>UT</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>06:47</td>
<td>2</td>
<td>00:02:92</td>
<td>UT</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>06:53</td>
<td>2</td>
<td>00:02:83</td>
<td>UT</td>
<td>Doubts about the figure shape</td>
</tr>
<tr>
<td>07:07</td>
<td>1</td>
<td>00:14:52</td>
<td>OM*</td>
<td>They asked for help</td>
</tr>
<tr>
<td>07:14</td>
<td>2</td>
<td>00:34:69</td>
<td>OM</td>
<td>They asked for help</td>
</tr>
<tr>
<td>07:52</td>
<td>1</td>
<td>00:39:95</td>
<td>OM*</td>
<td>They asked for help</td>
</tr>
<tr>
<td>07:57</td>
<td></td>
<td></td>
<td>Disassemble of part of the figure</td>
<td></td>
</tr>
<tr>
<td>08:05</td>
<td>2</td>
<td>00:32:56</td>
<td>OM*</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>09:05</td>
<td>2</td>
<td>00:15:88</td>
<td>OM*</td>
<td>Problems handling the pieces (2)</td>
</tr>
<tr>
<td>09:59</td>
<td>2</td>
<td>00:17:99</td>
<td>OM*</td>
<td>Problems handling the pieces (2)</td>
</tr>
<tr>
<td>10:31</td>
<td>1</td>
<td>10:41:22</td>
<td>NA</td>
<td>Problems handling the pieces (1)</td>
</tr>
<tr>
<td>10:36</td>
<td>2</td>
<td>00:02:76</td>
<td>UT</td>
<td>Problems handling the pieces (1)</td>
</tr>
<tr>
<td>11:04</td>
<td>2</td>
<td>00:02:44</td>
<td>UT</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>11:05</td>
<td>1</td>
<td>00:15:77</td>
<td>OM*</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>11:11</td>
<td>2</td>
<td>00:17:70</td>
<td>OM*</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>11:14</td>
<td>1</td>
<td>00:02:98</td>
<td>UT</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>12:10</td>
<td>2</td>
<td>00:03:43</td>
<td>UT</td>
<td>Problems handling the pieces (1)</td>
</tr>
<tr>
<td>12:18</td>
<td>1</td>
<td>00:02:49</td>
<td>UT</td>
<td>Problems handling the pieces (1)</td>
</tr>
<tr>
<td>12:54</td>
<td>1</td>
<td>00:02:50</td>
<td>UT</td>
<td>Problems handling the pieces (1)</td>
</tr>
<tr>
<td>13:10</td>
<td>1</td>
<td>00:02:26</td>
<td>UT</td>
<td>Problems handling the pieces (1)</td>
</tr>
<tr>
<td>13:30</td>
<td>1</td>
<td>00:02:65</td>
<td>UT</td>
<td>Problems handling the pieces (1)</td>
</tr>
</tbody>
</table>

Note. "*" = atypical behavior; "-" = no verbally expressed difficulty.

Table 3. Atypical behavior of Dyad 2

<table>
<thead>
<tr>
<th>Elapsed time</th>
<th>Participant</th>
<th>Duration</th>
<th>Atypical behavior</th>
<th>Related difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:00</td>
<td>1</td>
<td>01:44:28</td>
<td>NA</td>
<td>Not dialoging yet</td>
</tr>
<tr>
<td>03:16</td>
<td>2</td>
<td>00:54:85</td>
<td>OM*</td>
<td>Not dialoging yet</td>
</tr>
<tr>
<td>03:57</td>
<td>1</td>
<td>00:23:38</td>
<td>OM*</td>
<td>Not dialoging yet</td>
</tr>
<tr>
<td>04:41</td>
<td>1</td>
<td>01:49:64</td>
<td>NA</td>
<td>Not dialoging yet</td>
</tr>
<tr>
<td>05:18</td>
<td>2</td>
<td>00:03:30</td>
<td>UT</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>05:29</td>
<td>2</td>
<td>02:03:92</td>
<td>NA</td>
<td>Doubts about the course of action</td>
</tr>
<tr>
<td>06:39</td>
<td>1</td>
<td>02:35:58</td>
<td>NA</td>
<td>-</td>
</tr>
<tr>
<td>09:24</td>
<td>1</td>
<td>00:24:22</td>
<td>OM*</td>
<td>-</td>
</tr>
<tr>
<td>10:00</td>
<td>1</td>
<td>00:48:91</td>
<td>OM</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4. Atypical behavior of Dyad 3

<table>
<thead>
<tr>
<th>Elapsed time</th>
<th>Participant</th>
<th>Duration</th>
<th>Atypical behavior</th>
<th>Related difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:39</td>
<td>1</td>
<td>00:27.43</td>
<td>OM*</td>
<td>Not dialoging yet</td>
</tr>
<tr>
<td>01:04</td>
<td>2</td>
<td>05:09.18</td>
<td>NA</td>
<td>Not dialoging yet</td>
</tr>
<tr>
<td>01:52</td>
<td>2</td>
<td>01:33.72</td>
<td>OM*</td>
<td>Not dialoging yet</td>
</tr>
<tr>
<td>01:53</td>
<td>1</td>
<td>00:18.77</td>
<td>OM*</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>01:56</td>
<td>1</td>
<td>00:05.28</td>
<td>UT</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>02:26</td>
<td>1</td>
<td>01:06.10</td>
<td>NA</td>
<td>-</td>
</tr>
<tr>
<td>02:35</td>
<td>1</td>
<td>00:19.10</td>
<td>OM*</td>
<td>-</td>
</tr>
<tr>
<td>03:29</td>
<td>2</td>
<td>00:03.06</td>
<td>UT</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>03:36</td>
<td>2</td>
<td>00:02.84</td>
<td>UT</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>03:40</td>
<td>1</td>
<td>01:22.40</td>
<td>NA</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>03:49</td>
<td>1</td>
<td>00:24.35</td>
<td>OM*</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>04:00</td>
<td></td>
<td></td>
<td>Disassemble of part of the figure</td>
<td></td>
</tr>
<tr>
<td>04:31</td>
<td>1</td>
<td>00:18.14</td>
<td>OM*</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>04:32</td>
<td>1</td>
<td>00:04.88</td>
<td>UT</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>04:38</td>
<td></td>
<td></td>
<td>Disassemble of part of the figure</td>
<td></td>
</tr>
<tr>
<td>04:47</td>
<td>1</td>
<td>00:07.76</td>
<td>UT</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>04:56</td>
<td>1</td>
<td>00:27.52</td>
<td>OM*</td>
<td>-</td>
</tr>
<tr>
<td>05:05</td>
<td>1</td>
<td>00:05.30</td>
<td>UT</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>06:06</td>
<td>1</td>
<td>00:06.60</td>
<td>UT</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>06:16</td>
<td>1</td>
<td>00:15.98</td>
<td>OM*</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>06:16</td>
<td>1</td>
<td>00:08.54</td>
<td>UT</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>06:50</td>
<td>2</td>
<td>01:18.04</td>
<td>OM*</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>08:27</td>
<td>2</td>
<td>00:43.47</td>
<td>OM*</td>
<td>-</td>
</tr>
<tr>
<td>08:53</td>
<td>1</td>
<td>00:04.78</td>
<td>UT</td>
<td>-</td>
</tr>
<tr>
<td>08:54</td>
<td>1</td>
<td>00:25.10</td>
<td>OM*</td>
<td>-</td>
</tr>
<tr>
<td>08:59</td>
<td>1</td>
<td>00:06.10</td>
<td>UT</td>
<td>-</td>
</tr>
<tr>
<td>09:54</td>
<td>2</td>
<td>00:08.09</td>
<td>UT</td>
<td>-</td>
</tr>
<tr>
<td>10:51</td>
<td>1</td>
<td>00:04.95</td>
<td>UT</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>11:16</td>
<td>2</td>
<td>00:15.29</td>
<td>OM*</td>
<td>-</td>
</tr>
<tr>
<td>12:34</td>
<td>2</td>
<td>00:54.61</td>
<td>OM*</td>
<td>doubts about the pieces</td>
</tr>
<tr>
<td>12:35</td>
<td>2</td>
<td>00:03.29</td>
<td>UT</td>
<td>doubts about the pieces</td>
</tr>
<tr>
<td>12:58</td>
<td>1</td>
<td>00:40.37</td>
<td>NA</td>
<td>-</td>
</tr>
<tr>
<td>13:17</td>
<td>2</td>
<td>00:08.09</td>
<td>UT</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: "**" = atypical behavior; "-" = no verbally expressed difficulty.

Table 5. Atypical behavior of Dyad 4

<table>
<thead>
<tr>
<th>Elapsed time</th>
<th>Participant</th>
<th>Duration</th>
<th>Atypical behavior</th>
<th>Related difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:07</td>
<td>1</td>
<td>00:15.48</td>
<td>NA*</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>00:36</td>
<td>2</td>
<td>00:16.48</td>
<td>NA*</td>
<td>doubts about the figure shape</td>
</tr>
<tr>
<td>01:38</td>
<td>2</td>
<td>00:19.80</td>
<td>OM*</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>01:58</td>
<td>2</td>
<td>00:23.46</td>
<td>NA*</td>
<td>doubts about the course of action</td>
</tr>
<tr>
<td>02:06</td>
<td>1</td>
<td>00:23.79</td>
<td>OM*</td>
<td>doubts about the course of action</td>
</tr>
</tbody>
</table>

Note: "**" = atypical behavior; "-" = no verbally expressed difficulty.
The total number of atypical behaviors in the four dyads is 103. As mentioned, only one verbally expressed difficulty is not related to any atypical behavior; it occurred in Dyad 3 at 09:54 (see Table 4). On the contrary, in 22 of 103 occasions (21.4%), the atypical behavior was not related to any verbally expressed difficulty (those with a short dash in the Related difficulty column of Tables 2, 3, 4 and 5).

Longer utterances do not always follow long periods of inactivity as expected, but during the verbally expressed difficulty. Also, an expressed type of doubt always preceded the disassembling of part of the figure; this was also always related to an atypical behavior (see Table 2, 3, 4 and 5).

From the 103 atypical behaviors, were removed those that occurred: when the dyads had not established dialogue because there is no way to corroborate if they correspond to a difficulty (seven of them, those labeled as “Not dialoging yet” in the Related difficulty column in Tables 3 and 4); when Dyad 1 asked for help and was waiting for an answer (three of them in Table 2); and those of inactivity in object manipulation and navigation that did not correspond to inactivity in both (17 of them: those not marked with an asterisk “*” in the Atypical behavior column of Tables 2, 3, 4 and 5). In Table 6 are shown the results by dyad, where it can be seen the total of atypical behaviors presented (76, total of the second column) and the total of verbally expressed difficulties around the time that the atypical behavior took place (62, the total in the third column). In this same Table 6, the total of atypical behavior was break down by utterances (UT), inactivity in object manipulation (OM) and in navigation (NA) and their related difficulty when there was one. Difficulties were separated by doubts of any kind (probably more related to puzzlement and confusion), and problems handling the pieces (probably more related to frustration).

Regarding utterances, in Table 6 it can be seen that from 44 longest utterances: 29 are related to a kind of doubt and 11 to problems handling the pieces (the rest of them, five are not associated to a verbally expressed difficulty). No matter which participant had the problems handling the pieces, the longest utterances were made indistinctly by any of them (see Tables 2, 3, 4 and 5). The number of utterances by each participant was accounted to corroborate if the one who spoke the most was also the one with longest utterances, see Table 7; only in two dyads this was truth (Dyad 1 and Dyad 4).
### Table 7. Utterances by participant

<table>
<thead>
<tr>
<th>Dyad</th>
<th>Atypical utterances</th>
<th>Total of utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participant 1</td>
<td>Participant 2</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

About atypical inactivity in manipulation of objects (as can be seen in Table 6, the seventh column OM) it occurred 27 times: 18 related to different type of doubts, three to problems handling the pieces, and therefore six are not related to an expressed difficulty. Atypical navigation inactivity was present only in Dyad 4, five times: three related to a type of doubt, one to handling the pieces, so one is not related to an expressed difficulty.

The two dyads that had atypical behavior before they started the dialogue (Dyad 2 and Dyad 3) presented less expressed problems handling the pieces. The problems handling the pieces were at the end of the session, when the dyads were about to finish the figure, trying to fit the last pieces and managing the second lay of the figure.

In Table 8 are presented the amount of atypical behaviors from all the dyads, break down by type of difficulty.

### Table 8. Number of atypical behaviors in dyads by difficulty

<table>
<thead>
<tr>
<th>Atypical behavior</th>
<th>Doubts about the</th>
<th>Problems handling</th>
<th>Not related</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>figure shape</td>
<td>course of action</td>
<td>pieces</td>
</tr>
<tr>
<td>OM</td>
<td>3</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>NA</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>UT</td>
<td>13</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

**Evaluation by Dyad**

By doing a closer glimpse to each dyad data, it can be observed that in Dyad 1, only one atypical behavior of long utterances could not be explained with a verbally expressed difficulty (see Table 2 at 04:53). A number of long utterances are very representative when they had doubts about the figure shape, which ended in asking for help, and eventually in disassembling part of the figure.

In this particular dyad we found a peculiarity about their navigation. They navigated first by air, switched to earth navigation, by the end of the session Participant 1 switched again to air navigation. In Figure 4 are shown, the scatter-charts of navigation of this Dyad 1. The points connected the duration in navigation, a red circle for air navigation and a blue triangle for earth navigation. In the upper side of the figure is Participant 1 and in the bottom the Participant 2. Through the video, we realized that they talked about switching to earth navigation to facilitate the manipulation of the pieces.

For Dyad 2 (as shown in Table 3), a bigger number of difficulties seem not to be explained by any atypical behavior than as in other dyads. However, only the one at 09:24 represents inactivity in both object manipulation and navigation. This dyad presented a lack of communication in about the first half of the elapsed time of the session, as shown in Figure 5; they started to communicate and actually to collaborate by the end of the session. They also presented long periods of object manipulation inactivity compared with other dyads. Doubts about the course of action were mainly expressed with long utterances. This dyad was the one that ended doing the figure in a not correct way.

Dyad 3 (see Table 4) had long periods of inactivity at the beginning of the session while they were not talking to each other. This is the dyad with more atypical behavior not associated to any expressed difficulty; around a third of all they had (seven of 22). By the video it was appreciated that the participants decided to make one layer of the figure each one, and therefore they were not collaborating but having division of labor.
Figure 4. Earth and air navigation of Dyad 1

Figure 5. Utterances of both participants of Dyad 2
Dyad 4 was the only dyad that started almost immediately with a dialogue, and also they were the only one having atypical inactivity periods of navigation with inactivity in object manipulation, in both participants (see Table 5).

Conclusions and future work

We proposed the observation of nonverbal behavior during the accomplishment of a collaborative task in a CVE for the automatic understanding of affective behavior helpful for learning such as frustration, puzzlement or confusion in contrast to satisfaction or the determination to finish the task (Kort & Reilly, 2002). This is an indirect, not intrusive task-measure, but based on individual measures. As a novelty approach we decided to do an exploratory study on dyads.

Following Kantor (1921a; 1921b), we assumed an emotional no answer moment followed by its expression in the interaction with the others. In the exploratory study were collected inactivity atypical behaviors of the participants by measuring when they were not manipulating objects or navigating; also, unusual long utterances were collected. Afterwards the atypical behaviors were associated to verbally expressed difficulties (i.e., doubts about the course of action, the pieces to create the figure and the final shape of the figure to assemble, and also to problems handling the pieces). There is no way to comprehend the atypical behavior when the participants did not express themselves verbally, but that does not mean that they were not facing a frustrating, puzzlement or confusing moment.

Through the exploratory study we found, among others, a significant pattern, 62 times of 76 (a high rate for human behavior of 81.6%) atypical behaviors occurred around the time that the participants verbally expressed a type of difficulty related to frustration, puzzlement or confusing moments; although, neither long utterances nor inactivity could be associated to a particular type of difficulty. In a close observation to each specific case dyad, we found that in Dyad 3 they followed a division of labor approach, half of all dyads’ not related to a verbally expressed difficulties occurred in this dyad.

We are aware that the use of nonverbal behavior patterns is a holistic solution to understand affective behavior, and that this is just a particular situation for its application. In addition the small amount of trials, due to the fact that this is an exploratory study, does not allow making conclusive asseverations. However, the findings here encourage us to follow this path, exploring other nonverbal cues, as could be gaze direction or other pattern in utterances as the shorter ones. A probable way to better comprehend the affective behavior is to include during the utterances the collection of data related to voice inflexions. Also, the nonverbal behavior has been found helpful to explore, in the first place, if collaboration is actually taking place (Peña & de Antonio, 2009).

In a CVE, the collaboration phenomena can be easily collected and analyzed in time through nonverbal cues. This presents a significant resource to automatize affective computing and a way to automatically give feedback for apprentices/students or to aid a human or a virtual tutor to support learning. Some type of verbal analysis can complement this approach. The automatic analysis in affective learning should result in better scaffolding for the students in computer supported collaborative learning.

References


A Computational Model of Learners Achievement Emotions Using Control-Value Theory

Karla Muñoz1*, Julieta Noguez2, Luis Neri2, Paul Mc Kevitt1 and Tom Lunney1

1Ulster University, Magee campus, North Ireland // 2Tecnologico de Monterrey, Mexico City campus, Mexico // munoz_esquivel-k@email.ulster.ac.uk // jnoguez@itesm.mx // neri@itesm.mx // p.mckevitt@ulster.ac.uk // tf.lunney@ulster.ac.uk

*Corresponding author

ABSTRACT

Game-based Learning (GBL) environments make instruction flexible and interactive. Positive experiences depend on personalization. Student modelling has focused on affect. Three methods are used: (1) recognizing the physiological effects of emotion, (2) reasoning about emotion from its origin and (3) an approach combining 1 and 2. These have proven successful only in labs, or use theories of emotion not associated with an educational setting. The Control-value theory of achievement emotions holds that appraisals of control and value are most meaningful when determining emotion. This paper focuses on the design and evaluation of an emotional student model of Control-value theory applied to online GBL environments using Approach 2. This model is implemented using a dynamic sequence of Bayesian Networks (BNs). PlayPhysics - an emotional GBL environment for teaching Physics - was designed, implemented and evaluated with 118 students at ITESM-CCM. To evaluate our model, we employed cross-validation and Cohen’s Kappa. Our model achieved a fair to moderate accuracy of classification, but the results are promising. Future work will focus on identifying other variables that can improve classification.

Keywords

Affective student model, Game-based learning, Control-value theory, Dynamic sequence of Bayesian Networks

Introduction

Formal instruction is transforming into a more flexible and interactive process, focusing on student preferences for learning and engagement (Moore, Dickson-Deane, & Galyen, 2011). As a result, Virtual Learning (VL) and Game-Based Learning (GBL) environments have gained popularity and acceptance. GBL environments typically comprise features, such as storytelling, sound effects and feedback, which facilitate an emotional connection with the learner (Sykes, 2013). The key to attaining positive and successful experiences in VL and GBL environments is to achieve personalization (Janssen, van den Broek, & Westerink, 2011).

Emotion has shown to be important in many contexts including Evolution and Neuroscience. Here we focus on Educational Psychology and Computing contexts, where student modeling has recently focused on affect, because it has been shown to influence student understanding, performance and motivation. However, to date, the methods employed to reason about emotion in ITSs have shown highly promising in laboratories, but not in the classroom (Arroyo et al., 2009). For reasoning about emotion, the majority of the models use theories that have not originated from an educational setting (Conati & Maclaren, 2009; Jaques, Vicari, Pesty, & Martin, 2011; Landowska, 2013). Therefore, it may be possible that the targeted emotions do not actually occur during the teaching/learning experience. Also, the classification accuracy of these models is presented mainly using only percentages, so it is unclear that the effects are not random. On the other hand, the Control-value theory of achievement emotions by Pekrun, Frenzel, Goetz and Perry (2007), assumes that control and value appraisals are the most essential to determine emotion in an educational context. Achievement emotions are derived from performing activities and the pursuit of goals. Performance and achievement are judged against previously defined standards of quality. It was observed that this theory has not previously been utilized to create a computational model of student emotion. Therefore, here we focus on this objective.

For reasoning about student emotion, we employ a Cognitive-based Affective User Modeling approach, which allows applying what is known, in this case in the psychological educational field, to predict emotion (Martinho, Machado, & Paiva, 2000). Our model predominantly uses answers during in game dialogues and contextual variables, e.g., mouse location and the number of times help is asked, because it is mainly targeted at on-line GBL environments. Our model is implemented using a dynamic sequence of Bayesian Networks (BNs), since they can effectively manage the uncertainty of the domain and appropriately represent the temporal interdependencies. A preliminary
version of our model was discussed in Muñoz, Mc Kevitt, Lunney, Noguez and Neri (2013), the model presented here is the final result of performing further tests and employing more formal tools to conduct our analysis.

**Related work**

GBL environments, i.e., Edutainment, enhance learning by providing immediate feedback to student actions in simulated contexts. They have proven to attain student attention and engagement more easily than VL environments (Muñoz et al., 2009). Their success depends on the composition of diverse elements that gives them an emotional character (Sykes, 2013), such as penalizing errors and rewarding learning, e.g., through sounds, colors, narrative, and scoring. These elements combine to create a unique game-experience known as gameplay. Lazazzaro (2004) argues that this emotional experience is the source of the appeal of playing games. GBL environments can also be combined with ITSs to achieve adaptable and personalized instruction (Conati & Maclaren, 2009).

The new generation of ITSs aims to recognize and respond appropriately to student affect (Alexander, Sarrafzadeh & Hill, 2008; Conati & Maclaren, 2009; D’Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; D’Mello, Olney, Williams, & Hays, 2012; Jaques et al., 2011; Landowska, 2013; Porayska-Pomsta, Mavrikis, & Pain, 2008; Sabourin, Mott, & Lester, 2011). The main motivation for modeling emotions and moods arises from the field of Affective Computing (Picard, 1995). It has been noted that GUIs that do not consider student affect may impede and limit performance (Brave & Nass, 2008). Three main approaches are employed by the new generation of ITSs for recognizing or reasoning about student affect: (1) identifying physical and physiological effects of emotion, (2) reasoning about observable behavior from its origin, i.e., Cognitive-Based Affective User Modeling and (3) a hybrid approach combining both. Identifying the physical and physiological effects involves acquiring data related to student behavior using hardware, e.g., cameras, sensors and microphones. This data is processed and relevant features are selected and mapped to emotional states using opinions of judges or self-reports (D’Mello et al., 2008; Landowska, 2013; Sarrafzadeh, Alexander, Dadgostar, Fan, & Bigdeli, 2008). Processing this kind of data requires high bandwidth, which may deteriorate performance. The facial coding system by Ekman and Friesen (1978) is often used as a reference to map facial gestures to emotional states.

Reasoning about emotion from its origin, i.e., Cognitive-Based Affective User Modeling (Martinho et al., 2000), involves using cognitive psychology theories as a reference to reason about the elements that determine emotion. The most common theory employed using this approach is the OCC model (Ortony, Clore, & Collins, 1990). This theory has been adapted to be applied to the learning experience, because it was originally created to explain emotion in personal diaries. Therefore, it is possible that some of the emotions do not happen or do not occur in the described manner in educational settings. This approach can employ contextual variables related to student behavior, which are considered low bandwidth variables; as a result, it can be applied to diagnosis of emotion during on-line learning. This approach has shown promising results (Jaques & Vicari, 2007; Sabourin et al., 2011), but has not been as successful as the previous approach. A hybrid approach that combines both approaches is expected to be more successful than its constituent parts. However, it also inherits the weakness of its composite approaches (Conati & Maclaren, 2009). The hybrid approach involves acquiring data related to the student interaction (context), physical changes and physiological signals, and then it uses all this information in conjunction to a cognitive theory to determine student emotion.

Contextual variables have been successfully employed to determine student motivation (Del Soldato, 1993), self-efficacy (McQuiggan, Mott, & Lester, 2008) and goals and attitudes (Arroyo & Woolf, 2005). After reviewing Control-value theory (Pekrun et al., 2007), we noticed that these variables are also related to determining student emotion. Del Soldato (1993) uses variables, such as the number times the student asked for help, performance and the number of times the student quit, to determine student motivation. McQuiggan et al. (2008) employs intentional (e.g., number of problems solved), locational (e.g., current learning goal), physiological (e.g., heart rate) and temporal (e.g., time in current location) variables to classify student self-efficacy in the GBL environment Crystal Island. Arroyo and Woolf (2005) employed variables like the time invested per problem and the average number of hints given per problem to effectively determine student goals and attitudes. Also, it is observed that student gaze can be used to infer student attention (D’Mello et al., 2008; D’Mello et al., 2012). Rust (2010) show that the mouse position is also an alternative to infer a person’s concentration or attention.
Theoretical framework

Dynamic sequence of Bayesian networks for Affect modeling

For defining the structure of Bayesian models, it is necessary to know the conditional independent or dependent relations (CIDRs), which can be defined with assistance of a domain expert or obtained through statistical tests of historical domain data using a learning algorithm such as Peter-Clark (PC) or Necessary Path Condition (NPC). The chosen algorithm depends on the available amount of data to derive, train and evaluate the model. For defining the parameters of the Bayesian model, a learning algorithm such as Expectation Maximization (EM) can be applied to discrete chance nodes from observed data (Jensen & Nielsen, 2007). The selection of the evaluation method also depends on the quantity of data available. With a large quantity of data available for training and testing, a hold-out procedure (Bouckaert et al., 2012) can be employed. However, when data is scarce, a cross-validation approach is employed, i.e., the dataset is divided into $n$ sub-samples and one of these is held for testing the model and the $n-1$ sub-samples are employed for training. Probabilistic Relational Models (PRMs) can be used also to facilitate the derivation of Bayesian models. They are object representations of the domain (Sucar & Noguez, 2008). As a result, the domain is characterized as series of entities with properties and relationships between them (Koller, 1999).

Pekrun theory

Pekrun et al. (2007) proposes the Control-value theory to explain how emotion arises in educational settings. Control-value theory focuses on achievement emotions, which arise from activities and outcomes that are judged against standards of quality. This theory focuses on understanding when students feel in and out of control of relevant activities and outcomes. Control and value appraisals are the key cognitive elements employed to define achievement emotions. Control refers to student beliefs about their abilities, e.g., skills and strategies, to perform an activity and attain its goal. Value relates to the assigned value of the activity or the outcome from the student perspective, which can be focused on achieving success or avoiding failure, where success and failure have positive and negative connotations respectively.

Pekrun et al. (2007) argues that if one of the appraisals is lacking, there is no emotion. There are three kinds of achievement emotions: prospective-outcome, activity and retrospective-outcome emotions. Two dimensions are considered to define the type of emotion that a person is feeling: the object in focus (activity/outcome) and the time frame (during an activity or before/after an outcome). Table 1 shows the definition corresponding to activity emotions in terms of control and value appraisals. It was observed that this theory has not previously been used to create a computational model of student emotion.

<table>
<thead>
<tr>
<th>Object Focus</th>
<th>Value</th>
<th>Control</th>
<th>Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>Positive/Negative</td>
<td>High</td>
<td>Enjoyment</td>
</tr>
<tr>
<td></td>
<td>Positive/Negative</td>
<td>High</td>
<td>Anger</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>Low</td>
<td>Frustration</td>
</tr>
<tr>
<td></td>
<td>High/Low</td>
<td></td>
<td>Boredom</td>
</tr>
</tbody>
</table>

Experimental design

Goal

In this work, the proposed computational model of student achievement emotions considers the Control-value theory (Pekrun et al., 2007) as a reference.
Hypothesis

The hypothesis of this work is that an emotional student model, based in Control-value theory and using answers to questions in game-dialogues and contextual variables, will reason about student emotion non-randomly and accurately.

We decided to focus on diagnosing student emotion in on-line GBL environments, since we would have access to a larger student population. For reasoning about student emotion, we employ a Cognitive-Based Affective User Modeling approach, since it employs low bandwidth variables.

Recognition variables employed for reasoning about emotion

To select the recognition variables, we examined the Achievement Emotions Questionnaire (AEQ) by Pekrun, Goetz and Perry (2005) corresponding to emotions that arise before/during/after a lecture, which comprises motivational, cognitive, affective and physiological factors. After identifying the factors employed by Pekrun et al. (2005), which we summarized in Table 2, we decided to focus on the cognitive and motivational factors while diagnosing emotion in on-line GBL environments, because these can be inferred from the interaction and the context of the learning activity. The affective factors signaled by Control-value theory are considered as student self-report of emotion during game interaction.

Table 2. Summary of cognitive and motivational factors

<table>
<thead>
<tr>
<th>Before</th>
<th>During</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude towards subject/activity</td>
<td>Current attitude towards subject/activity</td>
<td>Past outcome/outcomes</td>
</tr>
<tr>
<td>Confidence beliefs towards probable outcome (self-efficacy)</td>
<td>Current level of confidence</td>
<td>Willingness to keep performing/mastering the activity (investing effort)</td>
</tr>
<tr>
<td>Attitude towards investing effort (subject/activity)</td>
<td>Current effort invested</td>
<td>Eagerness to make the outcome public</td>
</tr>
<tr>
<td>Prospective level of difficulty (subject/activity)</td>
<td>Perceived level of difficulty (subject/activity)</td>
<td>Resultant attitudes towards subject/activity</td>
</tr>
<tr>
<td>Internal/ external motivation to perform &amp; achieve an activity</td>
<td>Student Level of concentration</td>
<td>Internal/external attribution of the obtained outcome</td>
</tr>
<tr>
<td></td>
<td>Status of progress on fulfilling the activity goals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avoiding requesting or asking for help</td>
<td>Resultant confidence on own capacity/skills</td>
</tr>
</tbody>
</table>

For the prospective outcome emotions, corresponding to the time frame before performing the learning activity, we observed that student attitudes and beliefs related to the future performance, i.e., outcomes are key to determining these emotions. Therefore, we decided to enquire about this by using game-dialogues, while introducing the task and the game story. For the activity emotions, regarding the time frame while the student is interacting with the GBL environment, we decided to employ contextual variables that have proven to be significantly related to variables such as confidence, effort and self-efficacy in related work to diagnose student motivation (Del Soldato, 1993), self-efficacy (McQuiggan et al., 2008) and goals and attitudes (Arroyo & Woolf, 2005). We also decided to use as a basis the classification of variables, e.g., temporal, intentional, locational and physiological, proposed by McQuiggan et al. (2008) to define our contextual variables. However, our locational variables correspond to where student attention resides.

To diagnose retrospective-outcome emotions, we use the latest state of the variables presented in Table 3, in specific the latest outcome, independence (attribution of the final result), and the type of outcome (the willingness to keep interacting). We also include a new variable, publishing outcome, that is a variable related to the student intention to make the outcome public.
Table 3. Contextual variables for recognizing activity emotions

<table>
<thead>
<tr>
<th>Type of variable</th>
<th>Variable</th>
<th>Description</th>
<th>Associated factors to control or value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal</td>
<td>Interval of Interaction</td>
<td>The total time that the student has interacted, since the game challenge is started</td>
<td>Effort</td>
</tr>
<tr>
<td></td>
<td>Time to achieve learning goal(s)</td>
<td>The time that the student invested in achieving the learning goal the first time</td>
<td>Confidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Perceived level of difficulty</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Attitude towards the activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Concentration</td>
</tr>
<tr>
<td>Intentional</td>
<td>Outcome</td>
<td>The result that is most likely to be achieved and directly associated to student performance</td>
<td>Effort</td>
</tr>
<tr>
<td></td>
<td>Times asked help</td>
<td>The number of times that the student asked for help</td>
<td>Confidence</td>
</tr>
<tr>
<td></td>
<td>Attempts alone</td>
<td>The number of attempts by the student to solve the challenge alone (without help)</td>
<td>Perceived level of difficulty</td>
</tr>
<tr>
<td></td>
<td>Estimated independence</td>
<td>Results from the difference between the number of attempts alone and the number of times that the student asked for help</td>
<td>Attitude towards the activity</td>
</tr>
<tr>
<td></td>
<td>Overall attempts</td>
<td>The total number of student attempts with and without help</td>
<td>Concentration</td>
</tr>
<tr>
<td></td>
<td>Average quality of tutoring feedback</td>
<td>The average value calculated from the student qualitative evaluation related to how useful, he/she finds the help or instruction provided</td>
<td>Effort</td>
</tr>
<tr>
<td></td>
<td>Type of outcome</td>
<td>Indicates whether the student obtained a successful outcome, committed a misconception or quit the game challenge</td>
<td>Confidence</td>
</tr>
<tr>
<td>Locational</td>
<td>Focus coarse value</td>
<td>The average value of the mouse position on the screen associated to student location</td>
<td>Perceived level of difficulty</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Attitude towards the activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Concentration</td>
</tr>
</tbody>
</table>
Proposal for representing students achievement emotions

Through examining control-value theory (Pekrun et al., 2007), it was observed that there is mutual causation between antecedents and effects of achievement emotions over time. Also, control and value are defined as categorical variables in the Control-value theory. BNs and Binary and Multinomial Logistic Regression (BLR/MLR) can handle categorical variables appropriately. As a result, we decided to implement a dynamic sequence of BNs to represent student achievement emotions.

To define the dynamic sequence of BNs, we employed the methodology shown in Figure 1. We focused mainly on defining the BNs structure and learning their parameters. Once the design of game challenges is known and their elements are described in a class diagram, we employed Probabilistic Relational Models (PRMs) to define a preliminary and generic structure of our emotional student model.

Figure 2 shows the generic PRM derived and corresponding to control-value theory. However, in this structure relationships between causes and effects are not completely defined, see the activity emotions PRM in Figure 3, where all the relations are indicated as dotted lines, meaning they are uncertain.

Therefore, we will have to acquire data corresponding to the student interaction with the GBL environment, which we can analyze by employing Pearson correlations and BLR/MLR (gaining more insight about the variables that enhance the classification). Then, all the information acquired through Pearson correlations and BLR/MLR is used to solve uncertain relations and complete the definition of BN structures applying the necessary Path Condition (NPC) algorithm. Finally to define the BN parameters, we apply the Expectation Maximization (EM) learning algorithm using the collected data.

Figure 4 illustrates the concept of time frame $t-1$ in the context of our emotional student model. The interaction with the GBL environment may be visualized as a film tape, but its execution is not necessarily in sequence, since students’ actions define the order in which the elements of the GBL environment are accessed. Concurrently, each achievement emotions network serves a purpose in time. For example, the prospective-outcome emotions network is employed for reasoning about emotion in PlayPhysics’ game dialogues. The activity emotions network is employed for reasoning about emotions while students interact with the challenges in the GBL environment. Finally, the
The retrospective-outcome emotions network is used to reason about emotion in the instant of time that the outcome of the game challenge is presented to the student.

At the beginning, the student is presented with a game dialogue that introduces the game’s plot, the next game challenge and enquires about student beliefs/attitudes. At the end of each game dialogue students self-report their emotions and the prospective outcome emotions network can be employed at this moment to reason about emotion (see Figure 4(1)). Then students may proceed to interact with the game challenge, if students self report their emotion before completing the challenge or evaluate the feedback provided by the learning companion, the activity emotions network can be used at that instant of time for reasoning about emotion (Figure 4(3)). If the latest entry corresponds to the ongoing interaction with a game challenge immediately after the game dialogue, value $t-1$ and control $t-1$, the state of the contextual variables is evaluated using the activity-outcome emotions network (Figure 4(2)). It is also possible that the latest interaction corresponds to the event of notifying students of their outcome. However, since students can retry game challenges as many times as desired after receiving their result, value $t-1$ and control $t-1$ can may also come from the retrospective-outcome emotions network (Figure 4(4)). Finally, when students have been presented with their game outcome, they must self-report their emotion towards it and may decide to proceed with another challenge, thereby starting another game dialogue. In this case, value $t-1$ and control $t-1$ come from the retrospective-outcome emotions network (Figure 4(5)).

Figure 3. Activity emotions PRM

Figure 4. Time $t-1$ in our student model
Experiment design

*PlayPhysics*

To acquire data related to student interaction with a GBL environment and enable students to communicate their emotion over time, we created *PlayPhysics*, an emotional GBL environment for teaching Physics. *PlayPhysics* includes our emotional student model and will enable testing of the hypothesis of this work. However, *PlayPhysics* also has to assist students in learning Physics. This section discusses key aspects of *PlayPhysics*’ design. For looking at more detail see Muñoz et al. (2013).

![Figure 5. Game challenge and GUI of PlayPhysics](image)

*PlayPhysics* functional and non-functional requirements were defined by conducting an on-line-survey of a course in introductory physics from the Tecnologico de Monterrey, Mexico City campus (ITESM-CCM) and Trinity College, Dublin. As a result, *PlayPhysics* is focused on teaching the topics of Newton’s laws for particles and rigid bodies, Dynamics and Kinematics and vectors in 3D, which were judged as the most challenging topics.

*PlayPhysics* is a Role-Playing Game (RPG) and space adventure comprising challenges that must be overcome using knowledge of physics. The student is an astronaut with the mission of saving Captain Foster, who is trapped in space station Athena. The mission begins when the student is going to be launched from Earth to travel to Athena, which is located between Mars and Jupiter, and which is rotating with a constant acceleration. Each challenge is associated with one game level. Here, we focus on the first challenge.

An expert in Astrophysics from the ITESM-CCM assisted us in defining the game-sce- narios. The first game-challenge comprises the Alpha Centauri spaceship, which has been launched from the Earth, and the Athena station. Alpha Centauri is heading at constant speed towards Athena, see Figure 5. The main goal is that the student sets suitable values for physics variables of Alpha Centauri to stop along Athena’s rotational axis. However, to make this goal challenging, the student has to fulfill conditions such as defining a position that facilitates docking and entering to Athena before the fuel is exhausted.

**Subjects**

For our investigation, we invited students enrolled in a related Engineering undergraduate degree at ITESM-CCM, and in an age range between 18 and 23 years old. We acquired the data from 118 participants that interacted with *PlayPhysics*.
Objects

The first challenge is related to the topic of one-dimensional rectilinear motion. To achieve a successful outcome, constant deceleration has to be applied. Two constraint variables in this challenge are: (1) the initial distance \( D \) from Alpha Centauri to Athena and (2) the time remaining until fuel is exhausted \( T \). As a result, both variables are assigned randomly within specific value ranges: \( D \in [17, 50] \) m and \( T \in [80, 120] \) s, to make the solution of the challenge non-trivial. Students must concentrate on setting the values of the exploration variables.

The factors that must be considered to solve PlayPhysics’ first challenge appropriately are:

- Choosing the correct direction for the acceleration \( a \) of Alpha Centauri spaceship.
- Setting the magnitude of the acceleration of Alpha Centauri considering that humans black-out if \( a > 4g \), where \( g \) is the gravity acceleration at sea level \( (g = 9.81 \text{ m/s}^2) \).
- Not going beyond the fuel exhausting time, \( t_s \leq T \), and achieving the lowest relative error, \( e_d \leq 2\% \), in the breaking distance \( (d_s) \).

\[
e_d = \frac{d_s - D}{D} \times 100
\]

- Defining the lowest value for the breaking time \( (t_s) \).

These factors were implemented in PlayPhysics as rules to diagnose student knowledge. The simulation model is concerned with the representation of the physics domain.

Instrumentation

Students solved a pre-test, and afterwards interacted with the first challenge of PlayPhysics, and finally solved a post-test and qualitative questionnaire. Students self-reported their emotional state before, during and after performing the game activity.

During the interaction with the game challenge, the student’s emotion can be reported at any time, using the EmoReport wheel (Figure 6 (a)). The emotion relating to the outcome at the end of the challenge is always enquired (Figure 6 (b)), whether the challenge finishes due to an error or misunderstanding or due to a successful end. Learning companion M8- robot provides an emotional response every time the student reports his or her emotional state (See Figure 6).

Data collection, cleaning and analysis

We collected the data from 118 students at ITESM-CCM from the Faculty of Computing and Engineering, who interacted freely with PlayPhysics during one week. Through applying to the collected data NPC - in combination with the information obtained from applying BLR/MLR and Pearson correlations using SPSS - and EM algorithms using Hugin Lite, we obtained the dynamic sequence of BNs, comprised of the prospective-outcome, activity and
retrospective-outcome networks. The resultant activity emotions network is shown in Figure 7. We used 708 cases related to student game interaction to derive this BN, where 136, 122, 262 and 188 cases corresponded to anger, boredom, enjoyment and frustration. WEKA was employed to perform stratified random sampling in order to obtain 499 cases from the original 708 cases that we had, since we used the free version of Hugin Lite, which is limited to 50 states and 500 cases. Also, we employed WEKA to convert continuous to categorical variables using equal frequency binning to divide the variables into two or three categories, bearing in mind that control and value are also divided into three and two categories respectively.

We employed 10-fold cross-validation using the available data to determine the performance of each network over fresh data. The objective was to compare the performance of each network, and we obtained sensitivity, specificity, precision and accuracy measures of the networks (Han & Kamber, 2006).

True positives ($t_{pos}$) are positive tuples that were correctly labelled by the classifier. True negatives ($t_{neg}$) are negative tuples that were labelled by the classifier. False positives ($f_{pos}$) are negative tuples that were negatively labelled by the classifier. False negatives ($f_{neg}$) are positive tuples that were incorrectly labelled by the classifier.

Considering these definitions; it is possible to define sensitivity, specificity and precision from them. Sensitivity (ss) is the true positive ($t_{pos}$) recognition rate. Specificity (sp) is the true negative ($t_{neg}$) rate. Precision (prec.) is the percentage of tuples that actually belong to each labelled category. Accuracy (acc.) is a function of sensitivity and specificity. Results corresponding to the classification of the different types of achievement emotions are presented in Table 4-6.
From the prospective-outcome emotions, anxiety and hope are classified with 80% and 67.5% accuracy (see under *ss*). This is owed to classifying more appropriately control in the “Medium” category rather than the “High” category using answers to questions in game-dialogues.

### Table 4. Performance of the prospective-outcome emotions network

<table>
<thead>
<tr>
<th>Observed</th>
<th>Anticipatory joy</th>
<th>Anticipatory relief</th>
<th>Anxiety</th>
<th>Hope</th>
<th>sp</th>
<th>ss</th>
<th>prec.</th>
<th>acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anticipatory joy</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>0.857</td>
<td>0.300</td>
<td>0.375</td>
<td>0.567</td>
</tr>
<tr>
<td>Anticipatory relief</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>6</td>
<td>0.914</td>
<td>0.500</td>
<td>0.625</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0.925</td>
<td>0.800</td>
<td>0.571</td>
<td></td>
</tr>
<tr>
<td>Hope</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>27</td>
<td>0.660</td>
<td>0.675</td>
<td>0.613</td>
<td></td>
</tr>
</tbody>
</table>

*Note. sp = specificity; ss = sensitivity; prec. = precision; acc. = accuracy.*

From the activity emotions (Table 5), enjoyment and frustration are classified with accuracies (see under *ss*) of 67.8% and 60.0%, respectively. However, anger and boredom are classified with accuracies of 48% and 20%. This is due to being unable to recognize value appropriately in its category “None,” also the precision of frustration, is not very high, as a result, there is a high probability of classifying the other emotions as frustration.

### Table 5. Performance of the activity emotions network

<table>
<thead>
<tr>
<th>Observed</th>
<th>Anger</th>
<th>Boredom</th>
<th>Enjoyment</th>
<th>Frustration</th>
<th>sp</th>
<th>ss</th>
<th>prec.</th>
<th>acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>48</td>
<td>2</td>
<td>27</td>
<td>23</td>
<td>0.848</td>
<td>0.480</td>
<td>0.440</td>
<td>0.532</td>
</tr>
<tr>
<td>Boredom</td>
<td>15</td>
<td>18</td>
<td>31</td>
<td>26</td>
<td>0.954</td>
<td>0.200</td>
<td>0.486</td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>30</td>
<td>6</td>
<td>122</td>
<td>22</td>
<td>0.740</td>
<td>0.678</td>
<td>0.595</td>
<td></td>
</tr>
<tr>
<td>Frustration</td>
<td>16</td>
<td>11</td>
<td>25</td>
<td>78</td>
<td>0.808</td>
<td>0.600</td>
<td>0.523</td>
<td></td>
</tr>
</tbody>
</table>

*Note. sp = specificity; ss = sensitivity; prec. = precision; acc. = accuracy.*

### Table 6. Performance of the retrospective-outcome emotions network

<table>
<thead>
<tr>
<th>Observed</th>
<th>Anger</th>
<th>Gratitude</th>
<th>Joy</th>
<th>Pride</th>
<th>Sadness</th>
<th>Shame</th>
<th>sp</th>
<th>ss</th>
<th>prec.</th>
<th>acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>77</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>8</td>
<td>0.613</td>
<td>0.770</td>
<td>0.554</td>
<td>0.504</td>
</tr>
<tr>
<td>Gratitude</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0.932</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Joy</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0.992</td>
<td>0.100</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>Pride</td>
<td>11</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0.974</td>
<td>0.300</td>
<td>0.600</td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>22</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>8</td>
<td>0.880</td>
<td>0.450</td>
<td>0.529</td>
<td></td>
</tr>
<tr>
<td>Shame</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>16</td>
<td>0.918</td>
<td>0.400</td>
<td>0.471</td>
<td></td>
</tr>
</tbody>
</table>

*Note. sp = specificity; ss = sensitivity; prec. = precision; acc. = accuracy.*

On the other hand, from the retrospective-outcome emotions (Table 6), anger is classified with an accuracy of 77%. However, its precision it is not very high. Gratitude is not classified accurately at all.

### Table 7. Cohen’s Kappa for the achievement emotions networks

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Prospective-outcome emotions BN</th>
<th>Activity emotions BN</th>
<th>Retrospective outcome-emotions BN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kappa</td>
<td>Significance</td>
<td>Kappa</td>
</tr>
<tr>
<td>Emotion</td>
<td>0.369</td>
<td>7.732E-9</td>
<td>0.348</td>
</tr>
<tr>
<td>Value</td>
<td>0.550</td>
<td>1.811E-7</td>
<td>0.381</td>
</tr>
<tr>
<td>Control</td>
<td>0.311</td>
<td>0.003</td>
<td>0.429</td>
</tr>
</tbody>
</table>

Cohen’s Kappa (κ), an intra-class correlation coefficient, is employed as a measure of agreement that adjusts the observed proportional agreement by considering the amount of agreement expected by chance. Kappa can take values in a range [−1, 1], but only values in a range [0, 1] are meaningful, where the value of zero corresponds to
random classification. For hypothesis testing, if Kappa lies on the range: \(0.2 < \kappa \leq 0.4\), it corresponds to a fair agreement between the observed and the predicted values. If Kappa lies on the range: \(0.4 < \kappa \leq 0.6\), it corresponds to a moderate agreement. If Kappa lies on the range: \(0.6 < \kappa \leq 0.8\), it corresponds to a substantial agreement. The values of Cohen’s Kappa calculated for the achievement emotions networks are presented in Table 7. Control, value and emotion achieve fair-moderate classification accuracy and results are not random. This gives enough evidence to accept our alternative hypothesis.

**Evaluation and discussion**

*PlayPhysics* teaches physics and was created with the intention of providing instruction to students in an introductory course of physics. Therefore, *PlayPhysics* targets students in the last year of High school and first years of undergraduate education. In this work, we focused principally on assessing a student model of emotion for the target population using a Cognitive-Based approach. In this case, the Control-value theory by Pekrun et al. (2007) is the cognitive psychological theory used to derive the model.

Conati and Maclaren (2009) used the cognitive psychological theory of the OCC model (Ortony et al., 1990) as a reference for their model. However, this theory was not originally created to explain emotion in an educational context, but instead was created for reasoning about emotion in personal diaries. So, it is not clear if the emotions chosen are relevant to, or will arise in the same manner during the teaching-learning experience. Conati and Maclaren (2009) employed an Embodied Pedagogical Agent (EPA) to remind students to self-report their emotion. In a similar manner, *PlayPhysics* employs the learning companion M8- robot. Students can also use a pop-up window to report their emotion in Prime-Climb, and in similar manner, students using *PlayPhysics* can employ the *EmoReport* wheel. However, this is always present in *PlayPhysics*’ game challenges screen. Joy, distress, admiration and reproach are the emotions identified by PrimeClimb. Conati and Maclaren (2009), as other researchers, present the results corresponding to their emotional model using percentages of agreement between student self-reports and the predictions of the emotional model (69.59%, 62.30%, 67.42%, and 38.66% accuracy for joy, distress, admiration and reproach respectively), which makes it difficult to appreciate its reliability.

Sabourin et al. (2011) also focuses on recognising student achievement emotions using CRYSTAL ISLAND as does *PlayPhysics*. But, CRYSTAL ISLAND uses the appraisal based theory of learning emotions by Elliot and Pekrun (2007) as a reference. It differs from Control-value theory in that it relates the attainment of performance or mastery of goals and its valence with the experience of achievement emotions. In similar way, Sabourin et al. (2011) do not consider the category of “no-emotion” in their model, as in our investigation, since it is not defined by either theory. Their results are also reported as percentages of agreement, so it cannot be known whether the agreement is or is not random. Sabourin et al. (2011) focused on identifying student confusion, curiosity, excitement, focus, anxiety, boredom and frustration. The latter two were identified with accuracies of 18% and 28% respectively, whilst *PlayPhysics* identifies these two emotions using Control-value theory with accuracies of 20% and 60% respectively employing only contextual variables. Our emotional student model is the first and only model to date that was implemented using Control-value theory.

Another theory, adapted and employed to identify emotions in education using facial expressions is the theory by Ekman (1999), which has been successfully employed by Autotutor (D’Mello et al., 2008) in laboratories. However, this approach has still not proven effective in classrooms or on-line environments. Autotutor uses artificial neural networks to classify features of emotion. As a result, the emotional model is more like a black box and does not result in an intelligible model of emotion, i.e., does not provide further information about the participants or the affective domain. *PlayPhysics*’ emotional student model is intelligible and assists us in identifying factors that are considered actual predictors of control and value and the manner in which these are associated. The model also assists us in achieving an enhanced understanding of the student population.

We employed PRMs to achieve an enhanced understanding of the variables that may be considered while creating our emotional student model. Therefore, they facilitate defining Bayesian student models. This approach has been employed previously by Sucar and Noguez (2008), but for the purpose of defining a student model capable of identifying the level of a student knowledge or understanding. The application of the NPC algorithm for structural learning has been successfully employed in the area of telecommunications (Bashar, Parr, McClean, Scotney, &
Nauck, 2010) when scarce data is available. Here, we employ the same approach in combination with information acquired through applying BLR/MLR and Pearson correlations to solve uncertain relations. Pearson Correlations have been successfully employed as criteria for defining the structure of a Bayesian student model of attitudes (Arroyo & Woolf, 2005). We use the results of applying BLR/MLR as criteria for creating the network structure, since Bayesian models are a kind of Logistic Regression (Roos, Wettig, Grünwald, Myllymäki, & Tirri, 2005) and we can know the contribution of each selected variable to the prediction. We are not aware of any other research that employs BLR or MLR for this same purpose.

Conclusion and future work

We presented here an investigation about whether the creation of a computational model of student emotions using Control-value theory (Pekrun et al., 2007) can achieve a reasonable accuracy recognising student emotions in online GBL environments. PlayPhysics was implemented to test whether our emotional student model can be applied to GBL environments. Results showed that our model attains fair-moderate accuracy with results that are not random using answers in game dialogues and contextual variables. But, the resulting model is not highly accurate (Values of Cohen’s Kappa where $\kappa \geq 0.75$). Therefore, future work will focus on utilising other observable variables such as facial expressions, sentiment and speech to identify other features to enhance the classification of control and value. Also, the approach that we employed to derive the dynamic sequence of BBNs proved effective in creating an intelligible emotional student model and may be employed to derive other dynamic and intelligible data models to attain an enhanced understanding in areas other than education, e.g., e-Commerce and Genetics, in addition to the prospective areas of Affective Student Modelling and Adaptable Computer Tutoring.

References


Cross Space: The Exploration of SNS-Based Writing Activities in a Multimodal Learning Environment

Kwang-Soon Lee and Bong-Gyu Kim*
Department of English Language and Literature, Mokpo National University, Jeonnam, South Korea // leesue67@yahoo.com // bgk033@mokpo.ac.kr

*Corresponding author

ABSTRACT

This study explores the positive learning effect of formulating English sentences via Social Network Service (SNS; Kakao-Talk) on less proficient L2 university students’ (LPSs’) writing, when the application is utilized as a tool to link in and out-of class activities in a multimodal-learning environment. Its objective is also to promote LPSs’ participation in writing activities as well as to reduce their anxiety about potential mistakes through offering writing opportunities in a virtual space. The LPSs (N = 62) were regularly directed to construct their own sentences and opinions via Kakao-Talk in accordance with an instructor’s specific to-verb prototype after receiving the prepared materials via Kakao-Talk, which stimulates the “recalling process” on the previous class content. Qualitative and quantitative methodologies including correlation and corpus analysis were conducted. The results support the benefit of correlated factors: the efficiency of input, challenge, social presence, the usefulness of to-verb practice, and feedback. In particular, multiple-regression analysis presents a highly workable model to maximize the efficacy of learning. Multimodal Teaching and Learning with an SNS-based writing platform can have a significant effect in fostering LPSs’ writing performance through a specific grammar-driven (to-verb) approach, invoking learners’ motivation and enhancing peers’ collaboration as a newly-developed learning venue.

Keywords

Multimodal teaching and learning environment, SNS-based writing platform-Kakao-Talk, To-verb approach, Collaborative writing

Introduction

The innovative use of educational technologies has been developing at pace and offers higher education institutions valuable opportunities to design more inclusive media-enhanced learning environments. They also stimulate the online provision of off-campus study to convert traditional print-based materials into multimodal, interactive and technology-mediated e-Learning formats (Sankey, Birch, & Gardiner, 2010). Multimodal learning environments combine sensory modes (visual, aural, written) and content knowledge to improve learners’ attention and learning performance, especially for lower-achieving students (Moreno & Mayer, 2007). In particular, the educational elements of multimodal learning environments can mix other instructional systems such as e-Learning or m-Learning to promote learners’ engagement. Fadel and Lemke (2008) also uncovered the mixed and positive trends of multimodal designs combined with augmented learning. Students engaging in multimodal learning environments are reported to outperform those who learn in a traditional environment. Recognizing the positive features of multimodal learning environments, many scholars stipulate that learners should be using social media for such purposes, and that educators should also embrace these new platforms (Jenkins, 2009). Therefore, the combination of social-media and education is considered as the new way to create an enthusiastic teaching and learning atmosphere. This suggests the potential of using SNS as an educational platform and highlights the necessity of a transactional space that combines classroom-based education with enjoyable leisure activities for young people (Erstad, Gilje, & de Lange, 2007).

Literature review

The need of SNS (Kakao-Talk) as a writing tool

Writing not only has an interactive social-function to express writers’ thoughts and induce readers’ understanding (Widdowson, 2003), but also provides the opportunity for the learners to have control over the structure, grammar and content. The non-verbal language presents the possibility to select more appropriate English (Kim & Kim, 2006). However, despite its clear advantages as a communication mode, writing has some drawbacks with regard to regulating prompt feedback and error correction, due to the limited time and procedural problems (Shehadeh, 2011).
For L2-less English proficient students’ (LPSs’) effective writing, smartphones can be used as a learning agent to overcome the regulation of time and space, by utilizing “In-class content” in an SNS environment. A smartphone can lead to relevant opportunities for learners to be connected with both the content and the instructor, as it is the most popular and accessible device for communication (Dunlap, Furtak, & Tucker, 2009). Furthermore, it can increase self-esteem by building “social capital: mutual assistance and trust” through interpersonal relationships (Steinfield, Ellison, & Lampe, 2008) and collaborative group activities (Yamamura, 2011). Also, composition through a non-simultaneous board can reduce learners’ anxiety, developing more comfortable interactions than face-to-face writing (Mabrito, 2000).

These features implicate the positive effects of SNS-based writing practice on “affective domain.” In particular, “goal orientation,” “self-regulated strategy,” and “self-efficacy” are strongly related to academic satisfaction and achievement (Kim & Yan, 2014). Furthermore, by sharing peers’ ideas through online writing, learners can experience more interactive social processes than paper-based environments (Goldberg et al., 2003). For instance, with the use of Twitter, a friendly-learning environment can be crafted to enhance L2 learners’ confidence in their language performance (Antenos-Conforti, 2009) because of the convenience of editing, writing and revision.

Additionally, it’s noteworthy that the population of young adults using Kakao-Talk in Korea is ever increasing, unlike in the USA (Figure 1). Therefore, a multimodal learning environment combining mobile-based writing via Kakao-Talk is needed to promote learners’ motivation and involvement (Pollard, 2015).

The current state of Kakao-Talk use

Consumer Intelligence Research Partners (CIRP) conducted a survey about the 3 most frequently used mobile applications on 500 mobile phone-users from June to September, 2013 in USA. 45% of them used Facebook in Figure 1(a) (Sterling, 2013). This may be because Facebook is an open network with less stringent rules about membership, information disclosure, and interaction than Kakao-Talk (Ahn, 2011).

However, since Kakao-Talk was released during the introduction period of smartphones in South Korea, March, 2010 (see Figure 1b), its usage has dramatically grown because of “the Multiplier effect of human network,” and “first-mover advantage.” The diffusion speed of the Kakao-Talk network slowed and passed from growth to maturity early in 2013 when the penetration rate of smartphones was over 70% (Lee, 2014). Additionally, as the Kakao-Talk service is available via Tablet, ipot-Touch and PC, the chat-app was used by 93% of smartphone-owners in South Korea and had 152 million users worldwide as of October, 2014 (Carter, 2014).

![Figure 1. Current state of mobile-app use (a) most frequently used apps in USA (b) Kakao-Talk users in Korea (2010 - 2013)](image-url)
According to Lee and Choi (2012) in Figure 2, English education-related apps were most frequently used in Korea. Among them, however, writing-based activities accounted for only 4%. Therefore, it is necessary to guide Korean students’ English writing experience via SNS. The cultural tendency of Kakao-Talk use also implies that Kakao-Talk can be adopted as a writing practice platform for LPSs’ writing enhancement and diverse interactions.

Figure 2. Smartphone apps use for English education

SNS-based phrasal-verbs writing practice

Online-writing approaches can be found via various SNSs. Shih (2011) reported on the advantages of using Facebook for writing proficiency and peer feedback. Kim and Yoon (2014) combined Mobile café (Mocafe) with Kakao-Talk by using “Fun sentences,” and “Fun essays” for EFL writing enhancement in mobile-based blended learning. Jin (2015) also selected Band as a writing tool of SNSs to enhance learners’ appetite to solve various grammatical tasks collaboratively in a group. However, we learn from these studies that students may need a more specific grammar-driven writing approach such as “to-verb phrases” to produce sentences with confidence.

Palmer (1988) emphasized modeling the significance of verbs in a sentence and their importance in writing. Kang (2004) suggested that educators ought to teach the form and meaning of verbs with pictures while analyzing any errors. Accordingly, learners’ writing and reading abilities can be improved by applying the structure of verbs (Chung, 2010). Considering that verbs are usually activated alongside other elements such as adverbs, prepositions, nouns, infinitives or gerunds. Phrasal-Verbs (PVs) can be particularly useful in helping LPSs to recognize grammatical points and to pick up key expressions used by native speakers (Hsieh & Hsu, 2010).

With this linguistic and lexical knowledge, an SNS-based writing platform can promote the characteristics of m-Learning, the four Rs: Record, Relate, Recall, and Reinterpret (Low & O’Connell, 2006) by recalling learned content and personalizing it. This process helps activate students’ prior knowledge and language performance alongside the on-screen learning contents (Fadel & Lemke, 2008). Thus, a key element of the instructional model is to optimize learners’ online experience and develop “affective domain” enough to stimulate virtual interpersonal relationship.

Research questions

The research questions were designed according to 2 hypotheses.

(1) Does a Multimodal Learning Environment with SNS (Kakao-Talk)-based Writing Activities have a positive effect on LPSs’ writing improvement with to-verb facilitation?

(2) What effect does a Multimodal Learning Environment with SNS (Kakao-Talk)-based Writing Activities have on LPSs in an affective domain?

- Sense of Presence
- Sense of Challenge
- Sense of Efficacy
Methodology

Background of the study

Figure 3 and 4 summarize the steps of technology-integrated multimodal teaching and learning model (MTLM). It illustrates ways to connect face-to-face classroom learning with a technology-based learning mode beyond the traditional space (Picciano, 2009). In pre-step of MTLM, participants were divided into 7 groups, each of which comprised 8-12 students. In step 1, for effective content knowledge, re-organized text-materials with audiovisual features were used in class. In step 2, after in-class activities, not only newly-developed class materials but phrasal-verbs including to-verbs were chosen as a prototype uploaded on Kakao-Talk, to induce LPSs’ engagement and reduce their burden on composition. It was also designed to develop LPSs’ awareness of the utilization of to-verbs and sentence comprehension on specialized topics given in class through “data-connectivity.”

![Diagram showing the steps of MTLM](image)

Figure 3. MTLM with SNS-based writing activity (1)

However, step 2 has a few restricted rules. Due to a strong familiarity with Kakao-Talk, most Korean students can take advantages of its technology such as “copying & pasting practice,” “using communicative nonverbal-emoticon as emotional cues,” (Luor, Wu, Lu, & Tao, 2010) and “remixing practice,” including “selecting, cutting, attaching and compounding semiotic resources” by uploading and downloading files. The technological familiarity can be a potential drawback in a communication system (Greenhow & Robelia, 2009). Accordingly, to prevent fake-response, “remixing” was allowed whereas just “copying & pasting,” and “exchanging emoticon without any written information” were prohibited, since they can block the potential for interactive writing and increase a language impoverishment. After SNS-writing activities, in step 3, learners were given essay assignments on 2 topics; “Future-mate,” and “Plastic surgery.”

During the learning process, the teacher avoided direct corrections over students’ mistakes on Kakao-Talk to reduce learners’ psychological reluctance against writing. As instructors’ prompt corrections through Kakao-Talk can be counterproductive and demotivating for maintaining learners’ further written interactions (Pollard, 2015). In this regard, spellchecking over Kakao-Talk writing was not required for participants as “over-accuracy” can curb LPSs’ voluntary involvement. Instead of individual correction, the teacher provided collective feedback with visual records of learners’ errors and in-class activities (see step 4). The teacher highlighted where common mistakes occurred in
class and provided question-specific comments. Thus, the primary objective of MTLM is to establish more opportunities to participate in writing without feeling any burdens of grammatical accuracy.

Participants

This study consisted of 62 LPSs taking general English as a liberal art with a TOEIC score of 500 or below, at M-University in South Korea. The experiment was conducted for about 15 weeks, 3 hours per week.

Data collection and analyses

To analyze students’ writing attitudes and opinions about SNS-based writing platforms, a pre-test and two types of post-test with surveys were conducted for a quantitative analysis using SPSS 18.0 version; reliability, correlation,
and multiple-regression analysis. Additionally, corpus analysis was implemented to see how much participants’ usage of to-verbs was improved. The test-data before and after the experiment was collected from only 38 students participating in both pre-test and post-test because of various personal reasons among the concerned students, while survey-data was extracted from 62 students.

**Versatile input design for MTLM**

Fundamentally, text-oriented materials were reformed for in-class activities in order to stimulate learners’ motivation, providing the effect of authenticity (Ellis, 2003). These materials contained “situational conversations” to enforce learners’ sentence formulation by adding to-verbs. In Figure 5, the song-combined YouTube creates an exciting class atmosphere to induce learners’ attention and learning goal, while practicing “to-verb patterns-facilitated dialogue” in a pair-work and a group-work.

![Figure 5. Creating dynamic learning atmosphere](image)

YouTube-“Clean the room” (Perry, 2011), “Walking the dog” (Aubrey, 2010) “Resting on the beach” (ScarefaceR18, 2010)

Scaffolding input to direct students’ writing activities was uploaded onto Kakao-Talk for learners to download and respond to for a limited period (Figure 6). The participants constructed various to-verb patterns on Kakao-Talk after in-class activities. This promotes “Record” function, to store specific information and apply perceived linguistic items for learners’ own performance as a learner-centric pedagogical experience (Low & O’Connell, 2006).

![Figure 6. Integrating in-class and off-class activities](image)

Customer & Waiter Dialogue - e.g., [Teacher: What would you like to have for dinner?]
[SI-] I would like to have Salmon Salad.
[S2-] Would you like California Roll?

Download for a limited period
Figure 7 shows a reorganized article, circulated directly to LPSs both in class and after class through Kakao-Talk to accumulate knowledge for active involvement in discussion (Godwin-Jones, 2007). The students exchanged their thoughts about the article on Kakao-Talk, guiding some new members. This authentic material matches persons outside the classroom for academic concepts, creating prompt visual responses between peers. Thus, “Relate,” or “Relevance” attribute enables learners to enhance writing outcomes by expressing their ideas from realistic experiences (Fadel & Lemke, 2008), which promotes learners’ consciousness-raising and self-monitoring (Pollard, 2015).

Data evaluation design

One-type of pre-test and two types of post-tests were prepared and carried out for 30 minutes respectively. Type I was used for both pre-test and post-test to assess participants’ writing proficiency development. It consisted of 15 questions in which students were directed to describe some pictures of daily life activities and subsequently connect two related sentences by filling in an appropriate word in the blank. Type II consisted of 7 multiple-choice questions to evaluate participants’ perception of to-verbs and 18 open questions to allow students to express their thoughts on this (Total = 25). The type II post-test, which was similar to the pre-test pattern but modified, was designed to estimate students’ practical ability.
All the test questions except the multiple-choice questions were marked using Writing Level Guidelines derived from the Writing Assessment Criteria of SMU-Multimedia Assisted Test of English (SMU-MATE), which is categorized into 4 grades: Expert, Commanding (high, mid, low), Moderate (high, mid, low) and Rudimentary (Jeong, 2010). However, the criteria in this study were redesigned into 3 grades (Rudimentary/Moderate-Low/Moderate-Mid), which is considered to be most suitable for the LPSs participating in this experiment. This is because most of them have no experiences in an authorized writing test such as TOEIC writing, and the criteria of SMU-MATE was designed for Korean undergraduates.

Table 1. Assessment criteria of participants’ writing proficiency test

<table>
<thead>
<tr>
<th>Writing proficiency grade</th>
<th>Credit</th>
<th>Qualities of writing proficiency for LPSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate Level (Mid)</td>
<td>3</td>
<td>• making simple sentences and conveying the meaning.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• arranging the order of Subject &amp; Verb.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• conjugating basic tenses.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• expressing sentences through pictures.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• successfully utilizing to-verbs.</td>
</tr>
<tr>
<td>Moderate Level (Low)</td>
<td>2</td>
<td>• partially expressing the meaning of sentences despite spelling &amp; grammatical errors.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• filling Korean with gaps between the words or showing a strong connection with Korean in using vocabulary.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• not perceiving to-verbs</td>
</tr>
<tr>
<td>Rudimentary Level</td>
<td>1</td>
<td>• minimally communicating with difficulty without employing sentences due to deficient vocabulary &amp; knowledge of English sentence order.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• simply listing isolated words.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• having difficulty conveying proper meanings due to the spelling &amp; grammatical errors with the confusion of English sentence order.</td>
</tr>
</tbody>
</table>

Reliability evaluation of post-survey

In this study, a post-survey questionnaire consisting of 35 questions was conducted based on a 5-point likert scale; (SA: strongly agree = 5), (A: agree = 4), (N: neutral = 3), (D: disagree = 2), (SD: strongly disagree = 1) and categorized into 6 variables: efficacy, presence, challenge, the usefulness of input, to-verb and feedback. To assess the internal consistencies of questionnaires, reliability analysis was conducted.

The reliability coefficient (0.900) of the post-survey represents strong internal-consistency (Table 2). The six constructs in Table 3 show high values ranging from 0.823-0.950 with the exception for feedback (0.587). However, feedback shows high corrected item-total correlation (0.707), despite relatively fewer items when compared with other constructs. Additionally, Cronbach’s alpha of if item deleted of each construct is less than reliability coefficients alpha (0.900). Therefore, the post-survey proved to be significantly reliable.

Table 2. Reliability statistics on post-survey

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Cronbach’s alpha</th>
<th>N of constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.900</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3. Item-total statistics on post-survey

<table>
<thead>
<tr>
<th>Constructs</th>
<th>N of items</th>
<th>Cronbach’s alpha</th>
<th>Corrected item-total correlation</th>
<th>Cronbach’s alpha if item deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Usefulness of Input</td>
<td>5</td>
<td>.823</td>
<td>.626</td>
<td>.897</td>
</tr>
<tr>
<td>2. Sense of Challenge</td>
<td>6</td>
<td>.928</td>
<td>.796</td>
<td>.872</td>
</tr>
<tr>
<td>3. Sense of Presence</td>
<td>4</td>
<td>.845</td>
<td>.645</td>
<td>.897</td>
</tr>
<tr>
<td>4. Efficacy</td>
<td>12</td>
<td>.926</td>
<td>.921</td>
<td>.855</td>
</tr>
<tr>
<td>5. To-verb</td>
<td>5</td>
<td>.950</td>
<td>.712</td>
<td>.886</td>
</tr>
<tr>
<td>6. Feedback</td>
<td>3</td>
<td>.587</td>
<td>.707</td>
<td>.886</td>
</tr>
</tbody>
</table>
**Results**

Before the experiment, the pre-survey on participants’ learning behavior was conducted. Table 4 shows the participants preferred to study individually (63.6%) and used SNSs for more than 30 minutes a day (81.8%), indicating the potential emotional stress of making mistakes (24.2%) and hesitation to ask for help. It also reveals that participants usually experienced problems with vocabulary (33.3%) and sentence structure (34.8%) and that few had writing experience (10.6%).

<table>
<thead>
<tr>
<th></th>
<th>Learning conditions</th>
<th>Study hours after school per week</th>
<th>SNS usage</th>
<th>The reason for Writing difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Individual study</td>
<td>Online learning</td>
<td>Private school</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>42 (63.6%)</td>
<td>9 (13.6%)</td>
<td>2 (3%)</td>
<td>13 (19.6%)</td>
</tr>
<tr>
<td>2</td>
<td>Not at all</td>
<td>Less than 1 hour</td>
<td>1-2 hours</td>
<td>2-3 hours</td>
</tr>
<tr>
<td></td>
<td>29 (43.9%)</td>
<td>28 (42.4%)</td>
<td>6 (9.1%)</td>
<td>3 (3%)</td>
</tr>
<tr>
<td>3</td>
<td>More than 30 minutes</td>
<td>About 30 minutes</td>
<td>Less than 30 minutes</td>
<td>Others</td>
</tr>
<tr>
<td></td>
<td>54 (81.8%)</td>
<td>9 (13.6%)</td>
<td>3 (4.54%)</td>
<td>(0%)</td>
</tr>
<tr>
<td>4</td>
<td>Vocabulary knowledge</td>
<td>Sentence structure</td>
<td>Emotional stress of making mistakes</td>
<td>Lack of Writing experience</td>
</tr>
<tr>
<td></td>
<td>22 (33.3%)</td>
<td>23 (34.8%)</td>
<td>16 (24.2)</td>
<td>Others</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7 (10.6%)</td>
</tr>
</tbody>
</table>

In open-ended questions, students expressed their preference for *Kakao-Talk* (80.3%) as a mobile-based multimodal learning tool and a more youth-friendly device, when compared with *Facebook* (12.1%). The reasons for frequent use mentioned were features such as immediate message checking, convenient mutual communication by monitoring recipients and its synchronous online-video technology.

The results were gathered in terms of 2 research questions.

**RQ1: The positive effect of multimodal learning environments with SNS (Kakao-Talk)-based writing activities on LPSs’ writing improvement with to-verbs facilitation**

Table 5 shows 38 participants’ writing proficiency, on average, greatly improved in type I post-test ($M = 32.789, SD = 9.823$), and type II post-test ($M = 37.500, SD = 14.799$), compared with pre-test ($M = 18.236, SD = 11.114$). It also shows *Mean Difference* (-14.552) between type I pre-test and post-test ($p = .000 < .01$).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Std deviation</th>
<th>Std error mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>Pre-Test</td>
<td>18.2368</td>
<td>11.11472</td>
<td>1.80305</td>
</tr>
<tr>
<td></td>
<td>Post Type I</td>
<td>32.7895</td>
<td>9.82319</td>
<td>1.59353</td>
</tr>
<tr>
<td></td>
<td>Post Type II</td>
<td>37.5000</td>
<td>14.79911</td>
<td>2.40073</td>
</tr>
</tbody>
</table>

**Table 6. Equivalent-form reliability of post-test type I and type II**

<table>
<thead>
<tr>
<th></th>
<th>Post Type I</th>
<th>Post Type II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Type I</td>
<td>Pearson’s correlation coefficient</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>38</td>
</tr>
<tr>
<td>Post Type II</td>
<td>Pearson’s correlation coefficient</td>
<td>.740**</td>
</tr>
<tr>
<td></td>
<td>Sig.(2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>38</td>
</tr>
</tbody>
</table>

*Note. **p < .01.*
Table 6 shows a strong correlation ($r = .740, p = .000 < .01$) between post-test type I and type II, which strongly indicates an equivalent reliability. This result demonstrates the SNS-writing program is statistically significant enough to improve participants’ writing proficiency.

Next, to evaluate students’ writing development in using to-verbs in their essays, their written texts were stored and analyzed through the Range and Frequency tool for the extraction of target words, to-verbs. Table 7 shows that the total token of words was increased during the study period. The number of to-verbs was also increased significantly as seen in 1st (117 times) and 2nd writing (141 times).

<table>
<thead>
<tr>
<th>Table 7. Writing results 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Total token</td>
</tr>
<tr>
<td>Total type</td>
</tr>
<tr>
<td>To-verbs</td>
</tr>
</tbody>
</table>

From Table 8, it can be assumed that the students tried to use various types of to-verbs for constructing meaningful sentences.

<table>
<thead>
<tr>
<th>Table 8. Writing results 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Writing: To+ verb</td>
</tr>
<tr>
<td>work, hear, build, judge, encourage, make, limit, stress, live, attend, exercise, prepare, clean, trust, sympathize, fight, fix, continue, happen, solve, look, fight (117)</td>
</tr>
</tbody>
</table>

Figure 8 shows the aspect and effect of SNS-writing practice regarding to-verbs. To be specific, students could raise their awareness of to-verbs patterns ($V1 = 64.5\%$) with advanced comprehension on the sentence-structures ($V2 = 67.7\%, M = 3.74$). Moreover, the participants expressed increased confidence in composition ($V3 = 61.2\%, M = 3.61$), and less anxiety toward writing activities ($V4 = 62.9\%$). They even put more effort into applying and memorizing previously learned patterns ($V5 = 52.9\%$). Note that the percent of positive responses represents the sub-sum of SA and A (%).

The scatter-plot in Figure 9 represents the positive inter-relations between the pedagogical usefulness of to-verbs patterns and the challenge from SNS-based writing practice. The more writing activities using to-verbs that the students participated in, the more challenge they experienced during the experiment.

Figure 10(a) shows learners’ linguistic facilitation in reformulating sentences on a specific to-verb pattern (e.g., “be pleased to”) given by a teacher, which reveals “creative sentence-choice.” Figure 10(b) represents students’ different
expressions about a visual source, which promotes an attribute of “Reinterpret” (Low & O’Connell, 2006). Peers’ writing on screen enhances “learner-centric construction.”

Figure 9. Scatter-plot

![Figure 9. Scatter-plot](image)

Figure 10. Learner-centric construction (a) Creative sentence-choice (b) Reinterpreted context

(a) 
I’m pleased to announce that our candidate is elected by a absolute majority of votes, so that we will have a party tonight.

(b) 
I was pleased to solve the problem of statics that it was a challenge for me every time.

However, the students still made mistakes and had questions requiring the instructor to make individual responses and corrections with regard to confusing to-noun, to-infinitive and to-gerund patterns (e.g., […]want to job…], […]want to fulfilling…]). Thus, the instructor provided tips for sentence-reformulation to correct students’ misuse on spelling and tense.

Figure 11. The effect of feedback on students’ writing

Figure 11 shows the effect of the teacher’s support on cultivating students’ writing in both face-to-face learning situations and SNS-virtual space (F1 = 67.7%). The teacher’s writing guidance with non-directive correction on SNS was considered to be particularly effective (F2 = 66.2%) in supporting students’ emotional relief. Moreover,
participants experienced a self-instructional effect from peer’s ungrammaticality (F3 = 46.8%). Through this writing process, students can have opportunities to self-review their own writing while “recalling” on their own learned linguistic knowledge (Low & O’Connell, 2006), which provokes “memory retrieval” (Fadel & Lemke, 2008). From this perspective, peer interaction can be useful to increase participants’ awareness (Nystrand & Brandt, 1989) and to attract more attention to their language forms (Tarone & Liu, 1995), establishing students’ collective power to change from “individual novices” into “collectively competent learners” (Donato, 1994).

RQ2: Influence of SNS (Kakao-Talk)-based writing activity on LPSs’ affective domain: Presence, challenge, efficacy

Sense of presence

Riva (2009) defines the sense of presence as a mental status where one feels as though one is there in the moment when fully engaged in media use; the awareness of the co-presence. With this, learners can participate actively, experiencing significant outcomes. Therefore, the sense of presence can be a factor in making a learning experience successful. 53.3% of the participants expressed strong engagement in the SNS-based writing activity (P1), which could stimulate “a sense of belonging” to the group (Figure 12). Also, participants showed voluntary involvement in the writing activity (P4 = 54.8%) through synchronous visible online-writing. They built up social presence (SP) by establishing a close relationship with an instructor (P3 = 54.9%) and experienced collaborative learning using peers’ writing on SNS (P2 = 62.9%).

SNS-based writing shows an advantage over non-verbal communication through “emoticons” (Figure 13a and 13b). While constructing sentences, students could show personal anxiety about “mandatory military service” with non-
verbal responses (Figure 13a) as well as emotional determination like “In order to capture your dream.” instead of borrowing other appropriate words (Figure 13b). The students were involved in a mutual dialogue beyond simply making to-verb related sentences. Particularly, Figure 13(b) portrays the SP between a teacher and a student by praising and responding. The student also used the L1 translation technique to convey its exact meaning for other readers engaged in SNS-based communication.

**Sense of challenge**

Providing an achievable challenge in language acquisition is essential in producing a successful outcome. Table 9 shows that participants could face various types of challenges while engaged in both SNS-based writing activities and LMS use. For instance, students developed the challenge through Kakao-Talk-based short writing (71%) and LMS-based essay writing (56.4%). Students seemed to feel their writing proficiency improved in SNS-based writing activities (64.5%) when compared to the LMS-based activities (59.7%). Therefore, it can be assumed that the participants preferred shorter writings as opposed to longer ones. With virtual writing, the participants appeared to have more thinking time than in face-to-face writing to organize well-structured sentences with appropriate words (69.4%).

**Table 9. Sense of challenge**

<table>
<thead>
<tr>
<th></th>
<th>SA</th>
<th>A</th>
<th>N</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Providing challenges for writing through Kakao-Talk</td>
<td>9(14.5)</td>
<td>35(56.5)</td>
<td>15(24.2)</td>
<td>3(4.8)</td>
</tr>
<tr>
<td>2</td>
<td>Improving English writing ability through Kakao-Talk</td>
<td>8(12.9)</td>
<td>32(51.6)</td>
<td>19(30.6)</td>
<td>3(4.8)</td>
</tr>
<tr>
<td>3</td>
<td>Having more thinking time in writing on Kakao-Talk</td>
<td>13(21)</td>
<td>30(48.4)</td>
<td>17(27.4)</td>
<td>1(1.6)</td>
</tr>
<tr>
<td>4</td>
<td>Providing challenges for writing through LMS</td>
<td>8(12.9)</td>
<td>27(43.5)</td>
<td>22(35.5)</td>
<td>5(8.1)</td>
</tr>
<tr>
<td>5</td>
<td>Improving English writing ability through LMS</td>
<td>8(12.9)</td>
<td>29(46.8)</td>
<td>20(32.3)</td>
<td>5(8.1)</td>
</tr>
<tr>
<td>6</td>
<td>Having more thinking time in writing on LMS</td>
<td>9(14.5)</td>
<td>31(50)</td>
<td>19(30.6)</td>
<td>3(4.8)</td>
</tr>
</tbody>
</table>

The participants’ challenge increased positively thanks to the teacher’s input during the writing process (Table 10). Participants addressed the usefulness of the teacher’s scaffolding class material (62.9%) and the visualized learning contents (61.3%). SNS-based writing guidance and practice could also reduce “anxiety” about language learning (54.8%), and short-sentence construction (56.5%). As a result, it enhanced participants’ motivation (59.7%), stimulating challenge.

**Table 10. Usefulness of pedagogical input**

<table>
<thead>
<tr>
<th></th>
<th>SA</th>
<th>A</th>
<th>N</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Usefulness of teaching material in classroom</td>
<td>8(12.9)</td>
<td>31(50)</td>
<td>22(35.5)</td>
<td>1(1.6)</td>
</tr>
<tr>
<td>2</td>
<td>Reducing language-learning anxiety</td>
<td>8(12.9)</td>
<td>26(41.9)</td>
<td>23(37.1)</td>
<td>5(8.1)</td>
</tr>
<tr>
<td>3</td>
<td>Enhancing learning motivation</td>
<td>9(14.5)</td>
<td>28(45.2)</td>
<td>19(30.6)</td>
<td>6(9.7)</td>
</tr>
<tr>
<td>4</td>
<td>Usefulness of visualizing learning material</td>
<td>12(19.4)</td>
<td>26(41.9)</td>
<td>21(33.9)</td>
<td>2(3.2)</td>
</tr>
<tr>
<td>5</td>
<td>Reducing emotional anxiety through short-sentence writing guidance on Kakao-Talk</td>
<td>12(19.4)</td>
<td>33(57.1)</td>
<td>22(35.5)</td>
<td>4(6.45)</td>
</tr>
</tbody>
</table>

Additionally, the following affirmative opinions were recorded in open-ended questions:

“I could notice my friends and their specific characters next to their writing performance on Kakao-Talk, which attracted my attention. It was cool. I surely felt involved in the process voluntarily.”

“The professor first wrote sentences and meanings to make students use the prototype of grammar structure. Eventually it contributed to less anxiety.”

**Efficacy and correlation between factors**

Figure 14 indicates learners’ efficacy on writing proficiency. Students experienced actual writing competence working with their teacher’s complementary examples (E2 = 72.6%) as well as development in their writing
technique through using SNS and LMS as a writing tool (E6 = 51.5%). The students developed their ability to recall acquired knowledge, and experienced a sense of achievement from attaining set outcomes (E3 = 64.5%). Students felt confident in SNS-based short-sentence composition (E9 = 58.1%) and LMS-based essay construction (E10 = 59.7%), while developing the capacity to express their own ideas through writing-sources such as peers’ written forms (E3 = 54.5%). Their confidence in writing is tied to “increased interest” in online-based writing activities; SNS (E1 = 58.1%) and LMS (E5 = 61.3%), which enable students to achieve better writing competence on topic-based paragraph constructions (E7 = 46.8%). This was noticed when they expressed active involvement in further online-writing practice (E11 = 43.5%).

Learners’ general satisfaction about each factor can be seen in Table 11. It presents the positive effect of multimodal-focused learning environments on the development of LPSs’ writing proficiency, emotional stability for learning, linguistic challenge, linguistic input and social involvement.

Table 11. Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
<th>Std. error of mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaffolding Inputs</td>
<td>62</td>
<td>3.6774(73.5%)</td>
<td>.63901</td>
<td>.08115</td>
</tr>
<tr>
<td>Challenge</td>
<td>62</td>
<td>3.7312(74.6%)</td>
<td>.67505</td>
<td>.08573</td>
</tr>
<tr>
<td>Presence</td>
<td>62</td>
<td>3.5847(71.7%)</td>
<td>.76010</td>
<td>.09653</td>
</tr>
<tr>
<td>Efficacy</td>
<td>62</td>
<td>3.7124(74.2%)</td>
<td>.63549</td>
<td>.08071</td>
</tr>
<tr>
<td>Use of To verb</td>
<td>62</td>
<td>3.7000(74.0%)</td>
<td>.73485</td>
<td>.09333</td>
</tr>
<tr>
<td>Feedback</td>
<td>62</td>
<td>3.6989(73.9%)</td>
<td>.62349</td>
<td>.07918</td>
</tr>
</tbody>
</table>

Figure 4. The criteria of efficacy
Furthermore, Table 12 shows how each variable is inter-related in an SNS-based writing environment to enhance learners’ self efficacy and affective domains. It is proven that all the variables generally follow a positive direction and strongly correlate with each other. In particular, efficacy shows a significantly high correlation with others (e.g., efficacy-challenge: $r = 0.830$, efficacy-presentation: $r = 0.760$, efficacy-to-verb: $r = 0.745$, $p = .000 < .01$).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Efficacy</th>
<th>Challenge</th>
<th>Presence</th>
<th>To verb</th>
<th>Input</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Efficacy</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Challenge</td>
<td>.830**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Presence</td>
<td>.760**</td>
<td>.624**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 To verb</td>
<td>.745**</td>
<td>.646**</td>
<td>.420**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Input</td>
<td>.626**</td>
<td>.581**</td>
<td>.482**</td>
<td>.461**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6 Feedback</td>
<td>.698**</td>
<td>.571**</td>
<td>.447**</td>
<td>.720**</td>
<td>.509**</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. ** $p < .01$

**Figure 15.** Change of efficacy (1) (a) Influence of challenge (b) Influence of to-verb

**Figure 16.** Change of efficacy (2): (a) Influence of presence (b) Influence of feedback
The interaction between each variable and efficacy is also shown visually on the following graphs. Figure 15(a) represents the change of challenge and efficacy based on the respective mean scores of each section. Figure 15(b) displays the increase rate of efficacy according to the mean-change of the to-verb as well as the score-distribution of each section in the box-plot. Each box denotes 50% of the students who belong to each section. The upper bar of the box represents the top quarter of students, while the lower bar of the box shows the bottom quarter of students. Therefore, Figure 15(b) presents solid evidence that the participants acquiring higher-scored to-verb results generally had higher-scored efficacy.

Figure 16(a) is portraying the mean of error bar. It is notable that students with a presence score of more than total mean of presence (3.584) attained increased efficacy and achievement. Efficacy in Figure 16(b) and Figure 17 shows a gradual increase, which illustrates positive influence of feedback and input.

![Figure 17. Influence of input on efficacy](image)

**Multiple-regression analysis**

*Multiple-Regression Analysis* was conducted to identify how the dependent variable “efficacy” responded to a change in a predictor variable and to verify the most significant variable for LPSs’ learning efficacy.

Model 4 in Table 13 shows the highest value of $R^2$ ($R^2 = 0.865, p = .000$) and Adjusted $R^2$ ($R^2 = 0.856, p = .000$) compared with the previous models. According to the Adjusted $R^2$ ($R^2 = 0.856$) in Model 4, it is revealed that challenge, presence, to-verb and feedback account for $85.6\%$ of the Sum change of efficacy. The result is considered to be significant with ($F\text{-change} = 4.303, df1 = 1, df2 = 57, p = .000 < .01$).

<table>
<thead>
<tr>
<th>Model</th>
<th>$R$</th>
<th>$R$-square</th>
<th>Adjusted $R$-square</th>
<th>$R$-Square change</th>
<th>Change statistics</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$F$ change</td>
<td>$df1$</td>
</tr>
<tr>
<td>1</td>
<td>.830$^a$</td>
<td>.689</td>
<td>.684</td>
<td>.35738</td>
<td>132.879</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>.886$^b$</td>
<td>.785</td>
<td>.778</td>
<td>.29951</td>
<td>26.423</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>.925$^c$</td>
<td>.855</td>
<td>.848</td>
<td>.24799</td>
<td>28.061</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>.930$^d$</td>
<td>.865</td>
<td>.856</td>
<td>.24122</td>
<td>4.303</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. $a$ = (Predictor): (Constant), Challenge; $b$ = (Predictor): (Constant), Challenge, Presence; $c$ = (Predictor): (Constant), Challenge, Presence, To-verb; $d$ = (Predictor): (Constant), Challenge, Presence, To-verb, Feedback.

Table 14 shows that Model 4 ($F = 91.591, p = .000$) is significant and most appropriate as the *Multiple-Regression Model of efficacy*. 

72
Table 14. ANOVA e

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Regression</td>
<td>16.971</td>
<td>1</td>
<td>16.971</td>
<td>132.879</td>
<td>.000a</td>
</tr>
<tr>
<td>Residual</td>
<td>7.663</td>
<td>60</td>
<td>.128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24.634</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Regression</td>
<td>19.342</td>
<td>2</td>
<td>9.671</td>
<td>107.802</td>
<td>.000b</td>
</tr>
<tr>
<td>Residual</td>
<td>5.293</td>
<td>59</td>
<td>.090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24.634</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Regression</td>
<td>21.067</td>
<td>3</td>
<td>7.022</td>
<td>114.185</td>
<td>.000c</td>
</tr>
<tr>
<td>Residual</td>
<td>3.567</td>
<td>58</td>
<td>.062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24.634</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Regression</td>
<td>21.318</td>
<td>4</td>
<td>5.329</td>
<td>91.591</td>
<td>.000d</td>
</tr>
<tr>
<td>Residual</td>
<td>3.317</td>
<td>57</td>
<td>.058</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24.634</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. a = (Predictor): (Constant), Challenge; b = (Predictor): (Constant), Challenge, Presence; c = (Predictor): (Constant), Challenge, Presence, To-verb; d = (Predictor): (Constant), Challenge, Presence, To-verb, Feedback; e = Dependent Variable.

The major variables having a crucial influence on efficacy are challenge, presence, to-verb and feedback in the model constant (Table 15). Standardized coefficients (β-coefficients) show a positive relationship between variables. It means that positive variable changes also have a positive effect on efficacy increases.

Table 15. Coefficients a

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized coefficients</th>
<th>Standardized coefficients</th>
<th>Collinearity statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
</tr>
<tr>
<td>1 (Constant)</td>
<td>.797</td>
<td>.257</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>.781</td>
<td>.068</td>
<td>.830</td>
</tr>
<tr>
<td>2 (Constant)</td>
<td>.477</td>
<td>.224</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>.548</td>
<td>.073</td>
<td>.582</td>
</tr>
<tr>
<td>Presence</td>
<td>.332</td>
<td>.065</td>
<td>.397</td>
</tr>
<tr>
<td>3 (Constant)</td>
<td>.162</td>
<td>.195</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>.343</td>
<td>.072</td>
<td>.365</td>
</tr>
<tr>
<td>Presence</td>
<td>.324</td>
<td>.053</td>
<td>.387</td>
</tr>
<tr>
<td>To-verb</td>
<td>.300</td>
<td>.057</td>
<td>.347</td>
</tr>
<tr>
<td>4 (Constant)</td>
<td>-.003</td>
<td>.206</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>.329</td>
<td>.070</td>
<td>.349</td>
</tr>
<tr>
<td>Presence</td>
<td>.307</td>
<td>.053</td>
<td>.367</td>
</tr>
<tr>
<td>To-verb</td>
<td>.322</td>
<td>.067</td>
<td>.257</td>
</tr>
<tr>
<td>Feedback</td>
<td>.153</td>
<td>.074</td>
<td>.150</td>
</tr>
</tbody>
</table>

Note. a = Dependent variable: Efficacy.

The results ascertain that Model 4 is the most appropriate working model: challenge (B = .329, t = 4.702, p = .000 < .01), presence (B = .307, t = 5.836, p = .000 < .01), to-verb (B = .322, t = 3.333, p = .002 < .01), feedback (B = .153, t = 2.074, p = .043 < .05). Moreover, Table 19 shows (VIF = 2.215 < 10), which proves that variables are independent and there is no “multicollinearity” problem. With all the results, the following equation can be established to calculate the change of efficacy.

Y(Efficacy) = -.003 + 0.329*(Challenge) + 0.307*(Presence) + 0.322*(To-verb) + 0.153*(Feedback)

Therefore, efficacy increases 0.329 per the change of challenge, 0.307 per the change of presence, 0.322 per the change of to-verb, and 0.153 per the change of feedback. The most valuable contributing predictor on efficacy for this learning platform is challenge followed by presence, to verb and feedback. In conclusion, the efficacy of learning can improve significantly when each variable interplays in a MTLM.
Conclusions

Fundamentally, it is not surprising that a mobile-based social media platform can influence the way people communicate with each other and construct their learning process to meet their objectives. However, without interaction between technology and mobile-users, or without considering communication behaviors in systematic teaching and learning venues, it is hard to find the educational effect of technology itself (Ahn, 2011). Moreover, though “affective domain” is considered difficult to measure, it is important to explore the variables to develop language learners’ proficiency and pave the road to achieve their learning goal. In this regard, this study reveals two possibilities to solve the existing concerns with experimental verification.

First, the empirical results of the study demonstrate that a mobile-combined MTLM creates a highly effective learning environment for L2-LPSs’ writing improvement by establishing social domain; “high level of learning motivation and involvement.” Furthermore, it highlights pedagogical implications of an SNS-based writing activity, through meaningful outcomes of potential variables such as solidifying specific grammar patterns (to-verbs), building social presence, promoting challenge and developing higher efficacy ($R^2 = 0.856$).

Second, this study provides a replica of an integrated curriculum design and interactive learning process. During Kakao-Talk-based writing process, the students gradually engaged with peers, showing various “leaner-centric writing behaviors” such as revising sentences, adding words and monitoring their own or peers’ writing. In particular, the instructor’s ample feedback and scaffolding guidance served an important role in drawing students’ interest and changing their perception of writing. This process can foster collaborative writing (Zeinstejer, 2008) by using peers’ various viewpoints, and promote conceptual challenge by achieving “remembering-self” through “selective memory retrieval” on the previous texts (Bruner, 1994). Accordingly, it helps students develop a task-based approach through predicting outcomes and explaining their ideas to themselves as well as to peers (Fadel & Lemke, 2008). With the technology-combined activities, learners can face well-organized and systematic learning challenges, while facilitating a newly-developed conceptual knowledge with a teacher’s clear direction, specific grammar-centered approach, and extended offline writing activities through LMS. Thus, technology serves not simply as a tool but as a pedagogical function of assisting “teaching and learning” to stimulate learner’s affect.

In terms of synchronous learning environments, however, arranging smaller-sized groups and a controlled group may be recommended for time-effective communication and objective constructs comparison. Furthermore, studies on affective expressions in learners’ writing and cognitive development with more detailed criteria may be needed for a comprehensive assessment of learning outcomes. All in all, though, as designing an SNS-based learning model requires the roles of faculty such as updating, delivering, and revision of materials, teachers should keep themselves informed with respect to learners’ needs and educational technologies. Accordingly, it is necessary to explore more effective ways of using “cutting-edge-technology” for timid students with less voluntary involvement.

References


[ScarefaceR18]. (2010). Years around the Sun, miles away [Video file]. Retrieved from http://www.youtube.com/watch?v=s0aaJgPGTig


Identification of Action Units Related to Affective States in a Tutoring System for Mathematics

Gustavo Padrón-Rivera1, Genaro Rebolledo-Mendez1*, Pilar Pozos Parra2 and N. Sofia Huerta-Pacheco1

1Facultad de Estadística e Informática, Universidad Veracruzana, Mexico // 2Universidad Juárez Autónoma de Tabasco, Mexico // zS12020111@estudiantes.uv.mx // grebolledo@uv.mx // pilar.pozos@ujat.mx // nehuerta@uv.mx

*Corresponding author

ABSTRACT

Affect is an important element of the learning process both in the classroom and with educational technology. This paper presents analyses in relation to the identification of Action Units (AUs) related to affective states and their impact on learning with a tutoring system. To assess affect, a tool was devised to identify AUs on pictures of human faces. Action Units are combinations of individual facial muscles or groups of muscles that create facial expressions in association with affect. Pictures from a population of students were taken while using an intelligent tutoring system for mathematics in a secondary school in a suburban school in Veracruz, Mexico. The students were asked to interact with the tutoring system for 40 minutes and they were photographed with the tool at a rate of 1 picture every 5 seconds acquiring a dataset consisting of 16,800 photos. To achieve identification, the software analyzes individual pictures using Principal Component Analyses (PCA) and Euclidian distance. The tool developed to classify affective states shows 88.88% accuracy in the identification of AUs when matching the recognized AU to the Cohn-Kanade AU-Coded facial expression database. The analyses also elicited the most common AUs for the population and their association with learning with the intelligent tutoring system. These preliminary results shed light on the issues of affect in relation to learning mathematics with tutoring systems and pave the way for the implementation of coping strategies based on the automatic recognition of facial expressions.

Keywords

Affect, Intelligent tutoring system, Action units

Introduction

A facial expression is a fundamental component of non-verbal human communication. It not only conveys responses to external stimuli but also motivates action and adds meaning and richness to human experiences. Affective computing (Picard, 2000) is a branch of computer sciences aimed at endowing technology with mechanisms for the recognition of, understanding of and reaction to human emotions. Emotions encompass a number of psychological and physiological reactions used as cues by others in non-verbal communication. These cues are a fundamental part of human communication and provide the rudiments to develop technology to detect emotions automatically.

The recognition of emotions is an open problem in the computer sciences (El Kaliouby & Robinson, 2005). Emotions influence all human activities, and the learning context whether in the classroom or with educational technology is no exception. Previous works have employed technology to recognize emotions by processing facial expressions, speech and other physiological signals (Shen et al., 2009). Of particular relevance to this investigation is the work of Craig et al. (2004) as it provides indications of the importance of emotions in the learning process mediated by educational technology. Craig et al. (2004) distinguished between emotions expressed in non-classroom settings such as joy or sadness and the emotions displayed during the learning process mediated by educational technology. As a consequence, Craig et al. (2004) proposed that only a sub-set of emotions is present in the educational context.

Following this distinction and with the aim of distinguishing between the bigger set of emotions and those specific to educational technology, the term “affective states” is employed in this paper to refer to boredom, frustration, confusion and cognitive engagement or flow (Craig et al., 2004). Although the expression of affect by students is a normal part of their interaction with educational technology, they could be more inclined to stronger affective reactions in situations involving problem-solving activities. In these contexts, educational technology is designed to encourage the resolutions of mathematical problems accompanied by cycles of mistakes and recovering from these mistakes. During these cycles, students could experience one or all the affective states as the activities proposed by
the tutoring systems often offer challenging activities (D’Mello et al., 2010). To fully endow educational technology with affective computing capabilities it is necessary to develop tools to recognize, understand and react to students’ affect. The recognition of affective states has remained elusive and focused on physiological and psychological aspects related to affect. Previous research on affect has mainly been directed at the recognition problem because of the difficulties and ambiguities associated with the recognition of human emotions (Ocumpaugh et al., 2012). Some progress has been made, but a reliable, autonomous and context-independent solution to recognize affect during the use of educational technology does not exist. One of the problems with the development of such a tool is the context in which recognition takes place.

This paper assumes that cultural differences among students (Terzis et al., 2013) add difficulties to the development of a generalized affect detector. This paper deals with the problem of recognition of affective states, aiming at providing evidence that the employment of action units (conceived as universal facial reactions) could aid the recognition and understanding of students’ affective states based on facial movements. The findings presented in this paper also support the idea that students undergo affective trajectories and suggest these have impact on the students’ learning. This paper also offers pointers in relation to affective display in a Mexican secondary school but does not elaborate on how these affective displays compare to students with a different cultural background. The hypothesis guiding this investigation is that it is possible to reliably detect affect in a tutoring system considering action units. The work is organized as follows: First the background literature for this work is presented, followed by the methodology employed during the empirical work and the results of a set of analyses. Finally, the paper concludes by discussing the results of this paper and suggesting future directions.

Background

Previous research in the area of students’ affective reactions to educational technology suggests that students experience a very specific set of emotions: boredom, frustration, confusion, engaged concentration or flow and a neutral state (Craig et al., 2004). To recognize affect with educational technology, different approaches have been developed and one of the most common is the employment of human observers to make judgements about the affect they see on students’ faces (Ocumpaugh et al., 2012). However, this methodology could be subject to human error because of the process of observation. To cope with these problems, previous works have used agreement between pairs of observers applying Cohen’s Kappa formula (Cohen, 1960) to establish whether two observers significantly disagree on their observations of affect. Although human observers for the determination of affect during the learning process is widely employed (Ocumpaugh et al., 2012), it can be obtrusive and expensive as it requires the orchestration of different hardware and software elements that may overwhelm the student and create a complicated experimental setting (Zeng et al., 2009).

Other problems such as delays in annotation or observation bias due to students feeling watched by observers could potentially lead to inaccurate results (Ocumpaugh et al., 2012). Because of the potential for error during the identification of affect, it is desirable to develop tools to detect affect automatically. To this end, different techniques have been developed including algorithms to reduce the amount of information needed to process digital images of the human face (Wong et al., 2001) or the use of cameras for tracking both facial movements and body posture while students were learning using a tutoring system (D’Mello & Graesser, 2012). During the study reported by D’Mello and Graesser (2012) the camera recorded the participants’ postures and faces before and after interaction with the tutoring system and the students were asked to self-report their affect in 20-second videos taken from the tutorial session. If they were unsure of their state, they could rewind the video to the previous fixed point and replay the 20-second segment.

The problems encountered during the study were associated with the inaccuracy of the self-reports and the obtrusiveness of the experimental setting. Following the idea of capturing the students’ faces, the experiment reported in this paper is less obtrusive because it proposes the use of a single webcam to record the students’ faces and, in considering Ekman’s et al. (1992) Facial Action Coding System (FACS) and its 46 action units (AUs), identifies the affective states present during interaction with a tutoring system. According to Ekman et al. (1992), combinations of AUs can create hundreds of facial expressions, for example fear is expressed with a combination of 6 AUs: $AU1 + AU2 + AU4 + AU5 + AU20 + AU260$. Ekman et al. (1992) defined basic emotions in terms of AUs; however, very little has been done to define the action units present in learning with educational technology. The rationale for employing Ekman’s et al. (1992) framework is that it might allow more accurate, objective and
reliable measurement of facial muscles that might be easily associated with affective expression during the use of a tutoring system. Similar studies employing Ekman’s et al. (1992) framework have suggested combinations of AUs associated with learning (McDaniel et al., 2007). Other works report the employment of different types of algorithms for image analysis (e.g., Gabor filters) and classifiers (e.g., Hidden Markov Models) to identify sequences of AUs in terms of head and shoulder gestures (Baltrušaitis et al., 2011).

Methodology

Materials

The materials employed for this research project consist of a detector of action units, a tutoring system and the learning tests.

A detector of Action Units was developed for this research project. The detector has capabilities for (a) the detection of Action Units in human faces as recognized from a still picture; (b) the extraction of facial features from a still image using statistical techniques and (c) the comparison, using the Euclidean distance (Danielsson, 1980) as a metric, with images taken from a well-established database of human faces and its Action Unit components. The Cohn-Kanade database (Kanade et al., 2000) contains images of 97 subjects showing emotions; all the example faces in the database are labeled with the Action Units that they are composed of. To recognize Action Units from a still picture the detector compares two components taken from that picture: “mouth-chin” and “forehead-eyes” and determines the Action Unit that best fits the details extracted from the still picture. See Figure 1 for a flow diagram explaining this process.

![Figure 1. Detection pair of Action Units per human face image](image)

The detector collects individual photographs consisting of still pictures of individual students’ faces. These pictures are taken at a rate of 1 picture every 5 seconds. To test the accuracy of the tool, the Cohn-Kanade pictures were fed into the detector and the results were compared with the actual description on the Cohn-Kanade database. The results of comparisons showed an accuracy of 70%. When the detector was wrong (30% of the test set), the discrepancy was likely due to similarities between Action Units. For example, AU43 (eyes closure) and AU45 (blink) are very difficult to recognize. The same situation happens between AU25 (lips apart), AU26 (jaw drop) and AU1 (inner brow raiser) and AU2 (raised eyebrows).

A tutoring system called Scooter the Tutor (Baker et al., 2007) was used in this study. The tutoring system was developed to teach how to interpret data and to create scatter plots to represent it. The system uses artificial intelligence to detect whether a student is “gaming the system” (Baker et al., 2006). This concept is used to indicate
whether students’ actions within the system encourage the provision of help in order to pursue correct answers. The tutoring system helps students learn while minimizing their gaming behavior. Students who game have a second chance to do the exercises while suggesting gaming is a waste of time and therefore results in extra exercises (Baker et al., 2006). Problems involve a variety of domains; an example of a scenario that a student must solve is:

“Samantha is trying to find out what brand of dog food her dog food Champ likes best. Each day, she feeds him a different brand and sees how many bowls it eats. But then her mom says that maybe her dog just eats more on days when he exercises more. Please draw a scatterplot to show how many bowls the dog eats, given the dog’s level of exercise that day.”

There were two isomorphic tests employed immediately before and immediately after the experiment. To prevent bias from either the pre- or post-test being more difficult than the other, the tests were the same as in previous studies (see, for example, Baker et al., 2006). The tests are referred to as Form A and Form B. Students with an even list number were assigned Form B for the pre-test and students with an odd list number were assigned Form A. For the post-test the tests were switched. Individual learning gains were calculated with the formula:

\[
\text{Learning gain} = \frac{\text{Posttest} - \text{Pretest}}{1 - \text{Pretest}}
\]

Participants

The population of students for this experiment \((N = 128)\) came from a secondary school in Coatepec, Veracruz, Mexico. There were four participating classes each comprising an average of 32 students. The average age of the participants was 14 years. However, because the detector works by taking still pictures of individual students’ faces, only 11 students per class were considered totalling 44 students as there were only 11 webcams available simultaneously per class. There were 24 females and 20 males. To avoid bias, the assignment of students to computers with webcams was random by generating random numbers in the range of the class’ list and without repetition. All the activities during the experiment were held during the class schedule. Consent (Department of Health, Education and Welfare, 2014) was requested from the students’ legal tutors and from the head teacher before carrying out this experiment.

Procedure

The study followed a quasi-experimental procedure consisting of four stages: (1) an instructional presentation, (2) pre-test, (3) intelligent tutor interaction and (4) post-test. The instructional presentation consisted of a slide show delivered by one member of the research team. The presentation involved an introduction to the topic at hand (scatter plots) as well as all the necessary conceptual background about scatter plots for students to learn from the tutor (Baker et al., 2004). The duration of the instructional presentation was 20 minutes and was followed by a pre-test. The pre-test was an assessment of a scatter plot that the student was required to plot using the information provided in the allotted 25 minutes. The second stage consisted of allowing students to interact individually with the intelligent tutoring software (Baker et al., 2006). The interaction lasted for 80 minutes and was split into two days. The students used the tutoring system during the time of their mathematics class and the teacher was present throughout the four stages of the experiment. During the two days of their interaction with the tutoring system the teacher could assist students with questions arising from their interaction but the teacher had not seen the intelligent software before. Because the aim of the study was to capture students’ affect via a webcam, our methodology allowed the collection of still pictures without the need for observers or intrusion from interruptions, thus allowing the students to express affect in a more natural setting than other studies.

Data

The final database consists of scores for the pre- and post-tests, learning gains calculated following Equation 1 and a collection of 480 images per student taken with the collector. Because the image detector collects two Action Units per picture, the final dataset of pictures consists of pairs of Action Units per record. Images that were blurry, had
poor quality, were out-of-focus or were incomplete (some students did not complete the two sessions in the study) were excluded, leaving the final database with pictures from 34 students. Pairs of Action Units in the dataset consist of two components: the upper face Action Units (forehead-eyes) comprising AU1, AU2, AU4, AU7, AU43, AU64 and the lower face Action Units (mouth-chin) comprising AU12, AU25, AU26. There are 18 possible combinations of Action Units that were labeled to identify unique affective expressions. For the purposes of studying affective states in relation to Action Units Table 1 proposes a preliminary labeling based on combinations of Action Units (upper and lower part of the face) and the association with affective states found in the study reported by McDaniels et al. (2007). This table includes pairs of Action Units where the sign “+” signifies that this action is present and the sign “-” signifies the action is not present.

Table 1. Combinations of Action Units and their association with Affective States

<table>
<thead>
<tr>
<th>AU Combinations</th>
<th>Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label 1 (+AU1, +AU12)</td>
<td>Frustration</td>
</tr>
<tr>
<td>Label 2 (+AU1, +AU25)</td>
<td></td>
</tr>
<tr>
<td>Label 3 (+AU2, +AU26)</td>
<td></td>
</tr>
<tr>
<td>Label 4 (+AU2, +AU12)</td>
<td></td>
</tr>
<tr>
<td>Label 5 (+AU2, +AU25)</td>
<td>Frustration</td>
</tr>
<tr>
<td>Label 6 (+AU2, +AU26)</td>
<td></td>
</tr>
<tr>
<td>Label 7 (+AU4, -AU12)</td>
<td></td>
</tr>
<tr>
<td>Label 8 (+AU4, +AU25)</td>
<td></td>
</tr>
<tr>
<td>Label 9 (+AU4, +AU26)</td>
<td></td>
</tr>
<tr>
<td>Label 10 (+AU7, -AU12)</td>
<td></td>
</tr>
<tr>
<td>Label 11 (-AU7, +AU25)</td>
<td></td>
</tr>
<tr>
<td>Label 12 (-AU7, +AU26)</td>
<td></td>
</tr>
<tr>
<td>Label 13 (AU43+AU12)</td>
<td></td>
</tr>
<tr>
<td>Label 14 (AU43+AU25)</td>
<td></td>
</tr>
<tr>
<td>Label 15 (AU43+AU26)</td>
<td></td>
</tr>
<tr>
<td>Label 16 (AU64+AU12)</td>
<td></td>
</tr>
<tr>
<td>Label 17 (AU64+AU25)</td>
<td></td>
</tr>
<tr>
<td>Label 18 (AU64+AU26)</td>
<td></td>
</tr>
</tbody>
</table>

Note. *Denotes states reported by McDaniels et al. (2007) as being associated with the upper part of the face only.

Results

There were different analyses made on the data collected during the experiment. Table 2 shows the descriptive statistics for the learning gains. As in previous studies (Rodrigo et al., 2012) with the same tutoring system, the learning gains from pre- to post-test are significantly different ($t(34) = -9.5171, p < .001$) for the population, confirming previous results with this tutoring system.

Table 2. Descriptives for learning gains, pre- and post-test for the population

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning gains</td>
<td>34</td>
<td>0</td>
<td>1</td>
<td>0.588</td>
<td>0.32</td>
</tr>
<tr>
<td>Pre-test</td>
<td>34</td>
<td>0</td>
<td>0.5</td>
<td>0.083</td>
<td>0.14</td>
</tr>
<tr>
<td>Post-test</td>
<td>34</td>
<td>0</td>
<td>1</td>
<td>0.622</td>
<td>0.305</td>
</tr>
</tbody>
</table>

Table 3 shows the percentages associated with the 18 affective labels identified in our study. Please note that the labels actually occurring during the experiment are only 10 out of 18.

A dataset was created consisting of time series for every student. The time series consists of sequences of labels (see Table 1) as they were captured with the video camera during the interaction with the tutoring system (see Figure 2).

To analyze the transitions, the time series was used as input for the SAX (Symbolic Aggregate Approximation) algorithm (Lin et al., 2003). This algorithm is used for converting the original time series into a symbol string or word, as it is known in the context of SAX. The time series of length $n$ is reduced into another represented by a string...
with length \( w \), where \( w < n \). Figure 3 shows an example time series being converted to a discrete string using the SAX algorithm. Both the size and the dimension of the dataset were reduced from 480 elements into a string of 5 letters.

**Table 3. Prevalence of labels shown in percentages**

<table>
<thead>
<tr>
<th>Label</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>56.12%</td>
</tr>
<tr>
<td>1</td>
<td>22.95%</td>
</tr>
<tr>
<td>2</td>
<td>0.56%</td>
</tr>
<tr>
<td>4</td>
<td>0.01%</td>
</tr>
<tr>
<td>5</td>
<td>0.02%</td>
</tr>
<tr>
<td>10</td>
<td>15.65%</td>
</tr>
<tr>
<td>11</td>
<td>2.10%</td>
</tr>
<tr>
<td>13</td>
<td>1.57%</td>
</tr>
<tr>
<td>14</td>
<td>1.01%</td>
</tr>
<tr>
<td>15</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

**Figure 2.** Dataset of pairs of Action Units per student with their learning gains

**Figure 3.** A time series with 480 elements reduced into a string of five letters employing the SAX algorithm

Learning gains were associated with the time series as shown in Figure 2. In order to understand the role of particular facial expressions in the learning process a cut-off point based on the mean of 0.5 was defined to analyze which path or trajectory considering the SAX reduction yields better learning gains (see Table 4). The entire sample was split
into two groups: Group one consists of 23 series with a learning gain value over the cut-off point and Group two consists of 11 series with a learning value below the cut-off point.

<table>
<thead>
<tr>
<th>Table 4. Examples of discretized trajectories in relation to learning gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVER cut-off point trajectories</td>
</tr>
<tr>
<td>d, f, g, g, d</td>
</tr>
<tr>
<td>f, e, f, g, d</td>
</tr>
<tr>
<td>e, g, g, g, c</td>
</tr>
<tr>
<td>e, g, g, e, c</td>
</tr>
</tbody>
</table>

The dataset was used as a training set with the intention of defining a classifier to predict learning gains for students based on the Action Units read by the detector. The ID3 algorithm was employed to train the predictor. The results of classifying the trajectories with 10-fold cross-validation were: correctly classified instances: 21 = 61.76 % and incorrectly classified instances: 13 = 38.23%. See Table 5 for the confusion matrix.

<table>
<thead>
<tr>
<th>Table 5. Confusion matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Over</td>
</tr>
<tr>
<td>Under</td>
</tr>
</tbody>
</table>

In order to test the classifier, a new experiment was carried out on a new population consisting of 47 students, 25 females and 22 males and an average age of 13 years. 22 students were randomly selected as there were only 11 webcams available per session and there were two sessions. The same experimental procedure was applied. The data consisted of 18 students and 4 students were discarded because they did not complete the 2 days of the experiment. The cut-off point based on the mean was also 0.5. For this dataset 16 trajectories were associated with higher learning gains and 2 trajectories with lesser learning gains. The classification accuracy for the new dataset is 88.88 %. The confusion matrix is included in Table 6.

<table>
<thead>
<tr>
<th>Table 6. Confusion matrix of the second experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Over</td>
</tr>
<tr>
<td>Under</td>
</tr>
</tbody>
</table>

The frequencies of affective states for the dataset (n = 34) can be seen in Table 7. Table 1 presents a preliminary association between Label 1 and frustration and Label 10 and confusion. The results suggest that students mostly displayed faces associated with frustration, followed by faces preliminarily associated with confusion. The following most frequent set of faces are Label 11 with no specific association with affective states.

<table>
<thead>
<tr>
<th>Table 7. Percentage frequency combinations of Action Units (AUs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU combinations</td>
</tr>
<tr>
<td>Label 0 (blured images)</td>
</tr>
<tr>
<td>Label 1 (AU1+AU12)</td>
</tr>
<tr>
<td>Label 2 (AU1+AU25)</td>
</tr>
<tr>
<td>Label 5 (AU2+AU25)</td>
</tr>
<tr>
<td>Label 10 (AU7+AU12)</td>
</tr>
<tr>
<td>Label 11 (AU7+AU25)</td>
</tr>
<tr>
<td>Label 13 (AU43+AU12)</td>
</tr>
<tr>
<td>Label 14 (AU43+AU25)</td>
</tr>
<tr>
<td>Label 15 (AU43+AU26)</td>
</tr>
</tbody>
</table>

The percentages allow an understanding of the overall amount of time for different affective states. This, however, does not provide an order or its relation to learning gains. A more comprehensive analysis using Principal Component Analysis (PCA) was carried out to discover patterns of behavior in the data. For this analysis the original database was employed (n = 34). We split the dataset into 3 stages, each stage containing 160 labels accounting for 13.33 minutes of interaction with the tutoring system. PCA was used to find out which Labels accounted for most of
the variability in the data. For the three stages Labels 1 and 10 were the principal components. Table 8 groups the students according to the three stages, the two principal components and the average their learning gains.

Table 8. Students grouped according to Principal Components

<table>
<thead>
<tr>
<th>Label 1</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Label 10</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students 11, 13, 14, 23, 33</td>
<td>Students 1, 2, 3, 4, 6, 7</td>
<td>Students 1, 2, 3, 4, 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students 1, 2, 3, 4, 6, 7</td>
<td>Students 11, 13, 14, 23, 33</td>
<td>Students 14, 24, 25, 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average learning gain</td>
<td>0.5820</td>
<td>0.5820</td>
<td>0.340</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8583</td>
<td>0.8583</td>
<td>0.898</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stage 1. The results of the PCA analyses show that two components account for 98.21% of variability suggesting that Label 1 and Label 10 represent the variability for the whole sample during the first 13.33 minutes of interaction. This analysis associates groups of students displaying mostly either Label 1 or Label 10 (see Table 8). The differences between the students’ learning gains associated with Labels 1 (frustration) and 10 (confusion) are significant ($W = 3.5$, $p = 0.03888$).

Stage 2. The results of the PCA analyses show that two components account for 98.21% of variability suggesting that Label 1 and Label 10 represent the variability for the whole sample during the second stage of the interaction. This analysis associates groups of students displaying mostly either Label 1 or Label 10 (see Table 8). The differences between the students’ learning gains associated with Labels 1 (frustration) and 10 (confusion) are significant ($W = 3.5$, $p = 0.03888$).

Stage 3. The results of the PCA analyses show that two components account for 96.01% of variability suggesting that Label 1 and Label 10 represent the variability for the whole sample during the third stage of the interaction. The students and their associated learning gains can be seen in Table 8. The differences between the students’ learning gains associated with Labels 1 (frustration) and 10 (confusion) are significant ($W = 1.5$, $p = 0.03977$).

Conclusions

Preliminary efforts for the identification of affective states have led to the development of a tool used to categorize affective states based on Action Units. An initial analysis considered the classification of labels achieved initial results of 61.76% recognition accuracy. By applying this tool to a different population of students interacting with a tutoring system, the prediction accuracy was 88.88% showing a fair level of prediction of learning gains based on labels. The results also suggest that students’ trajectories in association with the labels reveal that students who consistently had more instances of Label 10 (confusion) during the interaction with the tutoring system had better learning gains than students who had more instances of Label 1 (frustration). This result confirms previous findings suggesting the importance of feeling confused, particularly in tutoring systems for mathematics (Lee et al., 2011). We also made a hypothetical association between Label 1 (+AU1, +AU12) and frustration; the results showed that Label 1 accounts for a large percentage of instances in the sample and it was associated with poorer learning gains, as has been proposed in theoretical models (Kort et al., 2001) and found in previous studies (Craig et al., 2004). Label 10 was associated with confusion being the second largest type of instance found in this study. Being confused over a large period of time has been associated with higher learning gains (Andres et al., 2014).

Confusion has been found to have different effects on students and the dynamics of confusion (D’Mello & Graesser, 2014) and its impact on learning gains could be better explored with the tool presented on this paper. Similarly, there is evidence suggesting frustration also has different effects on learning (Baker et al., 2010) which could be explored with this tool. Unlike previous studies, the results reported in this paper were achieved with a tool for the automatic recognition of Action Units based on still pictures. Because the results reported in this paper mirror similar results found in other studies, these findings are preliminary evidence of the effectiveness of this tool. Future studies will seek to validate the tool on larger populations focusing on the detection of frustration and looking to investigate the dynamics of confusion toward the development of a framework based on the automatic detection of frustration and confusion in tutoring systems. The identification of affective states is often difficult and subjective because each human observer tends to have his own judgment.
The main contribution of this paper is that it provides evidence that the recognition of labels consisting of pairs of Action Units can aid the automatic recognition of affective states, particularly confusion and frustration. Although the work reported in this paper elaborates on previous efforts to associate Action Units with affective states (McDaniel et al., 2007) there are other tools to automatically detect Action Units (Sénéchal et al., 2013). The results presented in this paper, however, seem to confirm research findings suggesting the importance of confusion, a form of cognitive disequilibrium (D’Mello & Graesser, 2014) to accrue higher learning gains and the fact that frustration might lead to poorer learning gains (Kort et al., 2001; Craig et al., 2004) during the use of a tutoring system for mathematics. Because some videos originating in this study were discarded as they were blurred or unreadable, we manually reviewed the video recordings. An observation made during this review process suggests that students in this population collaborate in solving problems by asking help from their fellow pupils. This observation confirms previous studies with the same tutoring system at the same school suggesting a lot of collaborative work consisting of engaging in interdependently paced work and conducting work away from their own computer (Ogan et al., 2012).

Acknowledgements

The authors of this paper acknowledge the financial support of the CONACYT. We thank the teachers and students at the “Escuela Secundaria General Ignacio de la Llave” in Coatepec, Veracruz, Mexico for their help with the experiments. We also thank Prof. Ryan Baker for allowing us to experiment with “Scooter the Tutor.”

References


Analyzing How Emotion Awareness Influences Students’ Motivation, Engagement, Self-Regulation and Learning Outcome

Marta Arguedas1*, Thanasis Daradoumis1,2 and Fatos Xhafa3

1Department of Computer Science, Multimedia and Telecommunications, Open University of Catalonia, Barcelona, Spain // 2Department of Cultural Technology and Communication, University of Aegean, Mytilini, Greece // 3Department of Languages and Computer Systems, Technical University of Catalonia, Barcelona, Spain // martaarg2@gmail.com // adaradoumis@uoc.edu // fatos.xhafa@gmail.com

*Corresponding author

ABSTRACT
Considering social and emotional competence in learning, emotion awareness aims to detect the emotions that students show during their learning interactions and make these emotions explicit to them. Being aware of their emotions, students become more conscious of their situation, what may prompt them to behavioral change. The main goal of this work is to analyze the effects of emotion awareness, supported by specific teaching strategies, on students’ motivation, engagement, self-regulation and learning outcome in long-term blended collaborative learning practices. A bilateral goal also involves an initial study that explores the way emotion awareness affects teacher’s attitude and feedback as well as the competencies that teachers need to have in order to achieve a positive change on students’ affective and cognitive state. To this end a quasi-experimental study was designed with high school students. The results of this study show that when students are aware of their emotions and guided by specific teaching strategies, their learning performance improves in relation to their motivation, engagement and self-regulation. Likewise, when teachers are conscious of students’ emotional state their attitude and feedback become more effective and timely.

Keywords
Emotion awareness, Affective feedback, Affective learning, Motivation, Engagement, Self-regulation and learning outcome

Introduction
Emotion awareness and affective feedback emerge as important factors that influence learning process and learners’ performance (Calvo & D’Mello, 2010; Feidakis et al., 2013). To foster effective learning, teachers employ a student-centered constructivist approach, involving different cognitive and collaborative learning strategies (Rosenshine, 1997; Daradoumis & Kordaki, 2011). The combination of all these four elements leads to an integrated framework that aims to improve students’ motivation, engagement and self-regulation, and ultimately students’ learning outcome and skills during their collaborative learning processes (see Figure 1).

![Figure 1. Key factors that lead to effective learning outcome and skills](image)

Among cognitive strategies, cognitive dissonance is the perception of incompatibility between two cognitions, which can be defined as any element of knowledge, including attitude, emotion, belief, or behavior (Pintrich et al., 1993). The cognitive dissonance strategy holds that contradicting cognitions serve as a driving force that compels the mind...
to acquire or invent new thoughts or beliefs, or to modify existing beliefs, in order to reduce the amount of dissonance (conflict) between cognitions (Aimer, 1998; Lee et al., 2003). According to Piaget’s theory (1967; 1980), when a child recognizes cognitive conflict (disequilibrium), this recognition motivates him or her to attempt to resolve the conflict and thus change his/her cognition, attitude, or behavior. Piaget called the process of resolving conflict “equilibration.” According to him, equilibration refers to the process of self-regulation that maintains a balance between “assimilation” and “accommodation.” Several educational interventions have been designed to foster dissonance in students by increasing their awareness of conflicts between prior beliefs and new information (e.g., by requiring students to defend prior beliefs) (Guzzetti et al., 1993). Moreover, according to Aronson (1995) and Graesser et al., (1996), creating and resolving cognitive dissonance can have a powerful impact on students’ motivation for learning. In our case-study, we use cognitive dissonance to increase motivation for learning as well as foster students’ inner self-regulation.

Among collaborative learning strategies, Jigsaw is a popular and extensively used strategy in which the members of the class are organized into “jigsaw” groups (Aronson & Patnoe, 2011; Perkins & Saris, 2001; Bratt, 2008). Students are then reorganized into “expert” groups containing one member from each jigsaw group. The members of the expert group work together to learn the material or solve the problem, then return to their “jigsaw” groups to share their learning. This process helps students improve their listening, communication, and problem-solving skills. In addition, the teacher’s role in the jigsaw is to facilitate learning and support students by encouraging them to help each other and to ensure that everyone in their group understands the material and will be confident presenting it to his/her group. Jigsaw strategy is an efficient way for students to become engaged in their learning, be individually accountable for their learning and achieve more self-regulation in their performance (DiDonato, 2013). This strategy maximizes interaction and establishes an atmosphere of cooperation and respect for other students.

Regarding emotion awareness, the main objectives of emotional education can be summarized as: gaining a better understanding of emotions and identifying the emotions of others (Pekrun, 2005); developing the ability to identify and control our own emotions (Goleman, 1995; Kort & Reilly, 2002); developing the ability to self-motivate and change negative emotions into positive (Gardner, 2006); and, managing conflict in a positive way (D’Mello et al., 2007; Baker et al., 2007). Consequently, the ability of students to perceive emotion was positively related to peer bonding (Han & Johnson, 2012), whereas empathy with the learner’s emotion would increase their motivation in learning (Pérez-Marín & Pascual-Nieto, 2013).

Based on emotion awareness information, the teacher can provide affective feedback ensuring in that way students’ emotional safety and their engagement or persistence in the learning experience (Feidakis et al., 2014). An adequate (timely, situation-aware and personal) affective feedback can cause a change in the students’ emotional state, which can redirect their focus of attention and can induce a change in the way they think, act and interact with others, as well as it can regulate their behavior in a learning situation (Shen et al., 2009; Bahreini et al., 2012). In this sense, it is also important to know the emotional competencies that teachers should have in order to provide the most adequate affective feedback to their students (Jennings, 2011). There are also specific types of affective feedback that use emotional reactions, such as applause, to reduce negative emotional states especially in male university students in specific educational situations such as computerized self-assessment testing (Liu et al., 2015).

In the recent years, research in emotion awareness in learning situations has focused on several issues that include: capturing the sentiments and the emotional states enclosed in textual information so that opinions and emotions embodied in them could play a key role in decision-making processes (Loia & Senatore, 2014); examining the impact of the so-called academic emotions (enjoyment, anxiety, pride, anger, hope, shame/fault, relief, boredom, hopelessness) on students’ ways of thinking and information processing (Pekrun et al., 2011); embedding emotion awareness into e-learning environments “ecologically,” by avoiding introducing obtrusiveness or invasiveness in the learning process (Feidakis et al., 2014); identifying patterns of emotional behavior by observing motor-behavioral activity (facial expressions, voice intonation, mouse movements, log files, sentiment analysis, etc.) (Heylen et al., 2005; Davis et al., 2008); using affect grid to measure emotions in software requirements engineering (Colomo-Palacios et al., 2011); studying motivation in work environments in the IT field (Sinha et al., 2014).

As discussed above, previous literature emphasizes the need of both emotion awareness and teacher’s affective feedback as two important elements in students’ learning process. However, there is not yet an extensive analysis of the relationship between emotion awareness and students’ motivation, engagement, self-regulation and learning outcome as well as emotion awareness and teacher’s attitude and feedback. The significance of this study is to bridge
this gap and provide a detailed analysis on the way emotion awareness affects students’ motivation, engagement, self-regulation and learning outcome if it is coupled with cognitive and collaborative learning strategies (such as cognitive dissonance and Jigsaw strategy respectively) which play an important role in reinforcing students’ motivation for learning, engagement and self-regulation. So far, no other study has conducted such an integrated analysis of all these key factors that lead to effective learning outcome and skills, as it is depicted in Figure 1. To this end, we first set our research goal, hypothesis and questions. Then, we present our case study description, and we explain how we address these questions through a real experiment with high school students. Next, we present the results of the experiment, we discuss the obtained results concerning the research questions set and we check the validity of our research hypothesis. Finally, we provide our conclusions with suggestions for future research.

Research hypothesis and aims

Goal: The main goal of this work is to analyze the effects of emotion awareness on students’ motivation, engagement, self-regulation and learning outcome in long-term blended collaborative learning practices. A bilateral goal also involves an initial study that explores the way emotion awareness affects teacher’s attitude and feedback as well as the competencies that teachers need to have in order to achieve a positive change on students’ affective and cognitive state.

Hypothesis: “Increasing the emotion awareness of learners, their learning outcomes improve in relation to their motivation, engagement and self-regulation. Besides, by increasing the emotion awareness of teachers, their attitude and feedback become more effective and timely.”

Independent Variable: \( X = \text{emotion awareness} \)

Dependent Variables:
- \( Y = \text{students’ motivation in learning} \)
- \( Z = \text{students’ engagement in learning} \)
- \( H = \text{students’ self-regulation} \)
- \( J = \text{learning outcome} \)
- \( K = \text{affective feedback} \)

Research questions

(1) Is there a significant correlation between students’ emotion awareness and their motivation and engagement in learning?

(2) Is there any significant correlation between students’ emotion awareness and their self-regulation and learning outcome?

(3) Is there any significant correlation between teacher’s awareness about students’ emotions and his/her attitude and feedback?

Methodology

Case study description

In this work, we were based on an emotion analysis model (Arguedas & Daradoumis, 2013) that integrates the four concepts mentioned before, that is, emotion awareness, affective feedback, cognitive strategies and collaborative learning strategies within an Activity Theory Framework (Engeström et al., 1999). This framework describes a problem based learning scenario where participants interact with learning objects by means of a specific cognitive strategy, such as cognitive dissonance, and a specific collaborative learning strategy, such as Jigsaw, in order to carry out goal-oriented activities.

In parallel, we also employed a discourse analysis method, based on the work of Arguedas et al. (2014), to analyze text and conversation generated by students collaboratively in order to identify and represent the students’ emotions that take place during these activities in a non-intrusive way. This information is shown to both teacher and students. This provides the teacher with the necessary emotion awareness in regard of the way students’ emotions appear and evolve over time, which enables him/her to offer students cognitive and affective feedback.
Participants

This experiment was carried out with a class of twenty-four fourth-year high school students, taking an introductory Computer Science course, using the Moodle platform. We divided students in six groups of 4 members and we chose three of these groups as the experimental group (EG) and the rest as the control group (CG). All students worked in a collaborative activity based on the Jigsaw strategy for 15 class sessions (5 weeks). All students (18 girls (75%) and 6 boys (25%)) had the same characteristics and background, and the election of CG and EG was done completely randomly. Students in the EG were informed for the emotions they experimented during the activity, so they were emotion aware all the time. In contrast, students in the CG had not any emotion awareness facilities.

At course presentation, all students were informed of their preferred learning style and their emotional intelligence level by having them take specific tests at the beginning of the activity. At the beginning of the activity, to acquire students’ learning styles, students answered a questionnaire based on VAK Learning Styles’ Questionnaire of Lynn O’Brien (1990). In addition, to measure the initial level of students’ emotional intelligence we used a questionnaire based on PEYDE’s Questionnaire (Gallego & Gallego, 2004).

As regards students’ learning styles, in the EG, 58% of them were visual, 25% auditory and 17% kinesthetic (Figure 2 (b)), whereas in the CG, 42% of the students were visual, 33% auditory and 25% kinesthetic as shown in Figure 2 (a).

Towards the levels of students’ emotional intelligence, in the EG the results were a 60.42% in Problem Solving Ability, 60.21% in Relationship Ability, 61.04% in Empathy, 59.79% in Emotional Control and 61.88% in Emotional Awareness (Figure 3 (b)). While in the CG, the results were a 60.00% in Problem Solving Ability, 60.00% in Relationship Ability, 61.46% in Empathy, 62.08% in Emotional Control and 61.67% in Emotional Awareness as shown in Figure 3 (a).
Procedure and data collection

The scenario included a collaborative learning activity which was implemented following the Problem-Based Learning method and the Jigsaw collaborative strategy. The topic of the activity was "Introduction to Internet" and it was carried out in the Moodle environment.

Based on the Jigsaw collaborative strategy, the learning activity was divided in ten stages which in turn were grouped around five tasks to facilitate their implementation as shown in Figure 4. For each task, the teacher provided all the necessary resources (documents and tools). Data was collected in the texts and dialogues produced by students during their group work. The teacher guided and gave support to the learning activity, by providing appropriate affective and cognitive feedback and encouraged the students to participate actively in building their knowledge.

To extract emotions from discourse created by students in virtual spaces (Wiki, chats and forum debates), we first used a sentiment analysis tool developed by Jurado and Rodriguez (2015). Then, we applied an extension of the Rhetorical Structure Theory (RST) tool (Arguedas et al., 2014) in order to obtain a graphical representation of the emotional structure of discourse. Through this, we managed to provide both teacher and students with means to be aware of the students’ emotions and their evolution over time in a non-intrusive way. In the end of the learning activity, we used a questionnaire that has been designed with both open-ended and closed-ended questions, taking our research hypothesis and questions into account.

More specifically, discourse (text and conversation) has been divided into segments which were analyzed in order to discover and show all the emotions that appear in them. With regard to Wiki text, division was carried out according to the intentional structure of the text (Grosz & Sidner, 1986); that is, each segment conveys a specific goal which is the result of the contribution issued by each group member. In regard of dialogue, divisions were carried out at two levels, first at the exchange level and then at the move level inside each exchange. Doing so, we created a clear association between the intentional and the emotional structure of discourse in both modes (text and dialogue). All segments were numbered sequentially and we refer to them as units of analysis. Both Wiki text and conversation were analyzed segment by segment by applying first Sentiment Analysis and then the extended RST. The obtained outcomes are displayed graphically, as shown in an example of a Wiki text in Figures 5 and 6.
1. **ACTIVITY:** Explain and document the subject “Introduction to Internet”

2. **TASKS ASSIGNMENT**
   (2.1) “First, our group has to establish the color codes that each of its members will use. Second, we have to split up in pairs and each couple will be in charge of two topics of the assigned activity, so we will manage to equally share the theoretical and practical parts. Let’s go boys!!”, said Cristina.
   (2.2) Cristina and Pedro—“We take care of the task *What is internet?* and of the task *Technical elements and how internet works*” “I choose red” said Cristina. “I choose blue” said Pedro.
   (2.3) Antonio and Pablo – “We take care of the task *Historic evolution of Internet* and of the task *The digital world: What services does Internet provide?*” “I choose orange!!” said Antonio “then I choose green!!” said Pablo.
   (2.4) “Planning is ready, let’s do this!” – Pablo says.

3. **TASK 1: What is internet?**
   (3.1.) “Internet is a global system of interconnected computer networks that use the Internet protocol suite (TCP/IP) to link several billion devices worldwide. It is a network of networks that consists of millions of private, public, academic, business, and government networks of local to global scope, linked by a broad array of electronic, wireless, and optical networking technologies.” – Pedro says.
   (3.2.) “Don’t forget to write down the source, Pedro!” – Cristina says
   (3.3.) “Oh! You’re right Cristina. https://en.wikipedia.org/wiki/Internet” – Pedro says

4. **TASK 2: “Historic evolution of Internet”**
   (4.1) “Here we have a timeline: http://www.w3.org/2005/01/timelines/timeline-2500x698.png” - Pablo says

5. **TASK 3: “Technical elements and how internet works”**
   (5.1) “This video explains how internet works and which are its technical elements: https://www.youtube.com/watch?v=j_LpdtKXPc” – Cristina says.
   (5.2) “Easy peasy! :P” – Pedro says

6. **TASK 4: “The digital world: What services does Internet provide?”**
   (6.1) “Internet provides us with services such as: Electronic mail (e-mail), world wide web, file transfer protocol (ftp), chat rooms, mailing lists, instant messaging, chats, news groups, remote access, file sharing, streaming media, internet telephony (VoIP), online gaming, etc.” – Antonio says (Various webs)
   (6.2) “Core digital technologies are evolving and converging rapidly, fueled by real-time, real-world data, driving us toward a Knowing Society and creating the foundation for an avalanche of innovative software platforms and other digital tools available and affordable to anybody and everybody, everywhere, virtually for any purpose” – Pablo says (https://www.eftonline.org/digitalworld2030.html)

7. “Well done guys! We finished! ☺” – Cristina says

*Figure 5. Text of wiki*
Figure 6. Emotions detected in the text (wiki) and graphical representation of the emotional structure of the text (wiki)
Let’s explain both Figures in more detail. Initially, a “give-information” exchange is initiated by move 1 and presents the activity topic. This exchange can be considered as successfully completed only when a final supporting move (move 7) is provided at the end. In this exchange, the predominant emotion is joy (happiness/satisfaction). In order to implement the activity, in segment 2, a group member (Cristina) initiates a “give-information” exchange with move 2.1 in which she explains how the task topics will be assigned and organized within the group. The exchange is completed by three consequent supporting moves (2.2, 2.3 and 2.4) contributed by the other group members. Here again, the predominant emotion is joy (happiness/satisfaction). Next, each pair of students works on its task.

The first task is initiated through a question in segment 3 in the form of an “elicit-information” exchange. This exchange is completed by three supporting moves (3.1, 3.2 and 3.3), contributed by both members of the pair. Here, the predominant emotion is anxiety expressed by one of the members (move 3.2), followed by shame expressed by the other member (in the beginning of move 3.3). The second task is set up as a problem in segment 4 in the form of an “ascertain-information” exchange and is resolved by one of the members of the pair (move 4.1), whereas the other member (move 4.2) confirms the given solution. Here, the predominant emotion is joy (happiness/satisfaction). Task 3 is again set as a problem (“ascertain-information” exchange) in segment 5 and is resolved as before (through moves 5.1 and 5.2). Here again, the predominant emotion is joy (happiness/satisfaction). Finally, task 4 is presented as a question in segment 6 in the form of an “elicit-information” exchange. Here, both members of the pair provide complementary answers (moves 6.1 and 6.2) which complete the exchange goal successfully. In this case, no obvious emotions are expressed by either member; so, their emotional behavior here is characterized as neutral.

Data analysis

Our goal was to obtain both quantitative and qualitative data in order to measure and evaluate learners’ emotional state concerning the following units of analysis:

- Emotion Awareness (EA)
- Affective Feedback (AF)

As regards EA, the questionnaire was composed of four questions that included 16 items classified into four categories, using a five-point Likert-type scale ranging from 1 (Almost never) to 5 (Almost always) and two open-ended questions requiring qualitative answer. As concerns AF, the questionnaire was composed of two questions that included 5 items, using a five-point Likert-type scale ranging from 1 (Almost never) to 5 (Almost always) and one open-ended question requiring qualitative answer. All the questionnaire items are shown in Figure 7.

<table>
<thead>
<tr>
<th>Labels</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.EA</td>
<td>Happiness/Satisfaction</td>
</tr>
<tr>
<td>4.2.EA</td>
<td>Sadness/Shame</td>
</tr>
<tr>
<td>4.3.EA</td>
<td>Fear/Anxiety</td>
</tr>
<tr>
<td>4.4.EA</td>
<td>Anger/Frustration</td>
</tr>
<tr>
<td>5.1.EA</td>
<td>Motivated</td>
</tr>
<tr>
<td>5.2.EA</td>
<td>Concentrated</td>
</tr>
<tr>
<td>5.3.EA</td>
<td>Unsafe</td>
</tr>
<tr>
<td>5.4.EA</td>
<td>Bored</td>
</tr>
<tr>
<td>6.1.EA</td>
<td>Student has shown Solidarity</td>
</tr>
<tr>
<td>6.2.EA</td>
<td>Student has given Suggestions/Opinions</td>
</tr>
<tr>
<td>6.3.EA</td>
<td>Student was in Opposition to</td>
</tr>
<tr>
<td>10.1.EA</td>
<td>Student self-regulates his participation in the activity on time</td>
</tr>
<tr>
<td>10.2.EA</td>
<td>Student changes her behavior (towards more positive) faster</td>
</tr>
<tr>
<td>10.3.EA</td>
<td>Student gets involved to create and share knowledge on time</td>
</tr>
<tr>
<td>10.4.EA</td>
<td>Student improves his performance before it’s too late</td>
</tr>
<tr>
<td>10.5.EA</td>
<td>Student lightens her workload</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Labels</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1.AF</td>
<td>Teacher has used dynamic methodologies that have motivated me to learn</td>
</tr>
<tr>
<td>8.2.AF</td>
<td>Teacher has attended my feelings and emotions when there was a conflict in the group</td>
</tr>
<tr>
<td>8.3.AF</td>
<td>Teacher has facilitated group discussion to manage emotions</td>
</tr>
<tr>
<td>8.4.AF</td>
<td>Teacher has encouraged and motivated my individual work, sharing it with the team</td>
</tr>
<tr>
<td>8.5.AF</td>
<td>Teacher has resolved my questions and offered advice and suggestions</td>
</tr>
</tbody>
</table>

Figure 7. The questionnaire items
Regarding the statistical techniques employed in the analysis of the questionnaire data, we used descriptive statistics, calculating relative frequencies (%), as well as graphics to represent reality objectively. We also used bivariate correlation and analysis of variance to find relationships between the variables under study for each of the questions of our study.

**Results**

**Reliability statistics**

To ensure the reliability of data collection, we applied the Cronbach’s alpha coefficient as well as the skewness and kurtosis for each variable that was examined in order to check for multivariate normality. In this sense, the absolute values of skewness and the absolute values of kurtosis did not exceed a univariate skewness of 2.0 and a univariate kurtosis for each variable that was examined in order to check for multivariate normality. In this sense, the absolute values of skewness and the absolute values of kurtosis did not exceed a univariate skewness of 2.0 and a univariate kurtosis of 7.0; as such, we assumed that there was no critical problem regarding multivariate normality. Finally, the results of descriptive statistics obtained are described as follows:

Cronbach’s alpha is considered to be a coefficient of reliability (or consistency). A reliability coefficient of .70 or higher is considered “acceptable” in most social science research situations. All values of Cronbach’s alpha in Table 1 are higher than .70, which reinforces the reliability of our test scores.

<p>| Table 1. The Cronbach’s alpha coefficient and descriptive statistics of EA and AF, in CG and EG |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|</p>
<table>
<thead>
<tr>
<th>Cronbach’s alpha coefficient (EA)</th>
<th>Cronbach’s alpha coefficient (AF)</th>
<th>Cronbach’s alpha coefficient (EA)</th>
<th>Cronbach’s alpha coefficient (AF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.55</td>
<td>1.269</td>
<td>-0.276</td>
</tr>
<tr>
<td>SD</td>
<td>0.743</td>
<td>1.000</td>
<td>0.123</td>
</tr>
<tr>
<td>Variance</td>
<td>1.126</td>
<td>0.124</td>
<td>0.375</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.767</td>
<td>0.202</td>
<td>0.125</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.697</td>
<td>0.751</td>
<td>0.202</td>
</tr>
<tr>
<td>CA(®)</td>
<td>0.720</td>
<td>0.771</td>
<td>0.962</td>
</tr>
<tr>
<td>Emotion awareness (EA)</td>
<td>Affective feedback (AF)</td>
<td>Emotion awareness (EA)</td>
<td>Affective feedback (AF)</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------</td>
<td>--------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Mean</td>
<td>3.55</td>
<td>1.269</td>
<td>-0.276</td>
</tr>
<tr>
<td>SD</td>
<td>0.743</td>
<td>1.000</td>
<td>0.123</td>
</tr>
<tr>
<td>Variance</td>
<td>1.126</td>
<td>0.124</td>
<td>0.375</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.767</td>
<td>0.202</td>
<td>0.125</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.697</td>
<td>0.751</td>
<td>0.202</td>
</tr>
<tr>
<td>CA(®)</td>
<td>0.720</td>
<td>0.771</td>
<td>0.962</td>
</tr>
</tbody>
</table>

*Note. CA(®): Cronbach’s alpha if element is deleted.*

The values obtained from the descriptive statistics performed are described as follows:

With regard to Emotion Awareness (EA):
• Students’ Emotions (4.1.EA – 4.4.EA): The mean exceeded the value of three (3.0) in the items 4.1.EA (Happiness/Satisfaction), obtaining the values 3.55 in the CG and 3.58 in the EG.
• Students’ Mental States (Motivation) (5.1.EA - 5.4.EA): The mean exceeded the value of three (3.0) in the following items: 5.1.EA.-Motivation (CG 3.43 – EG 3.70) and 5.2.EA.-Concentration (CG 3.35 – EG 3.42).
• Students’ Behaviors (Engagement) (6.1.EA – 6.3.EA): The mean exceeded the value of three (3.0) in the following items: 6.1.EA-Solidarity (CG 3.32 – EG 3.57) and 6.2.EA-Provide Suggestions (CG 3.25 – EG 3.64).
• Attitude changes experienced by students (self-regulation skills) (10.1.EA – 10.5.EA): The mean exceeded the value of three (3.0) in all items in both groups; however, all item values in EG are higher than the ones in CG.
• From these results, at first glance EG students experienced higher mental states, behaviors and attitude changes than CG students. This indicates that Emotion Awareness, supported by specific teaching strategies, is strongly related to students’ motivation, engagement and self-regulation.

With regard to Affective Feedback (8.1.AF – 8.5.AF):
• The mean exceeded the value of three (3.0) in the items 8.1.AF (3.13), 8.4.AF (3.70) and 8.5.AF (3.70) in the CG and in the items 8.1.AF (3.32), 8.2.AF (3.28), 8.4.AF (3.88) and 8.5.EA (3.77) in the EG.

Here, we see that EG students benefited more from teacher’s attitude and affective feedback than CG students did.

Pearson’s correlations

Once we gathered the data obtained in the questionnaires, we calculated the Pearson correlation coefficient for the different variables to answer our research questions.

Regarding the first research question. We have correlated EA (X) with students’ motivation in learning (Y) and students’ engagement in learning (Z). To this end, we have used the data gathered in items 4, 5 and 6. The results obtained are shown in Table 2 and are presented graphically in Figures 8 and 9.

Table 2. Pearson correlation coefficient (n = 60) for research question 1

<table>
<thead>
<tr>
<th></th>
<th>CG</th>
<th>EG</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.EA</td>
<td>5.1.EA</td>
<td>.215*</td>
</tr>
<tr>
<td></td>
<td>5.2.EA</td>
<td>-.033</td>
</tr>
<tr>
<td>4.2.EA</td>
<td>5.3.EA</td>
<td>.275*</td>
</tr>
<tr>
<td></td>
<td>5.4.EA</td>
<td>-.213</td>
</tr>
<tr>
<td>4.3.EA</td>
<td>5.5.EA</td>
<td>.285*</td>
</tr>
<tr>
<td></td>
<td>5.6.EA</td>
<td>.080</td>
</tr>
<tr>
<td>4.4.EA</td>
<td>5.7.EA</td>
<td>.107</td>
</tr>
<tr>
<td></td>
<td>5.8.EA</td>
<td>-.003</td>
</tr>
<tr>
<td>Motivation</td>
<td>6.1.EA</td>
<td>.490**</td>
</tr>
<tr>
<td></td>
<td>6.2.EA</td>
<td>-.320*</td>
</tr>
<tr>
<td></td>
<td>6.3.EA</td>
<td>-.267*</td>
</tr>
<tr>
<td>Engagement</td>
<td>6.4.EA</td>
<td>-.110</td>
</tr>
<tr>
<td></td>
<td>6.5.EA</td>
<td>.561**</td>
</tr>
<tr>
<td></td>
<td>6.6.EA</td>
<td>-.069</td>
</tr>
<tr>
<td></td>
<td>6.7.EA</td>
<td>.093</td>
</tr>
<tr>
<td></td>
<td>6.8.EA</td>
<td>.400</td>
</tr>
<tr>
<td></td>
<td>6.9.EA</td>
<td>-.144</td>
</tr>
<tr>
<td></td>
<td>7.0.EA</td>
<td>-.003</td>
</tr>
</tbody>
</table>

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

As regards motivation, in the EG, taking into account that students had EA, when they felt joy (happiness), they were highly motivated (83.33%). However, when they experienced sadness they were bored (71.67%), but at the same time they tried to maintain their motivation to work. As concerns fear/anxiety and anger/frustration, these emotions led them to an unsafe/uncertainty mood (58.33%). In the CG, taking into account that students had no EA, when they experienced joy, they were motivated (70.00%) in the task. However, when emotions such as sadness and fear/anxiety appeared, they felt unsafe (70.00%) and bored (80.00%), whereas they showed a high tendency to opposition and disagreement, thus losing interest to continue developing their activities. In the case of experiencing anger or frustration, students of the CG felt unsafe (86.67%), i.e., they had a strong lack of self-confidence.

As regards engagement, in the EG, when students experienced joy, they had a supportive behavior (85.00%) with peers. When they experienced sadness, they made suggestions and gave their opinion (83.33%) to their peers. Fear and anger did not present any correlation. In the CG, when students experienced joy they maintained a supportive behavior - solidarity (76.67%) with peers. But when they felt sadness and fear, they showed low solidarity as well as opposition (58.33%) to the suggestions of their peers.
Regarding the second research question. We have correlated EA (X) with students’ self-regulation (H) and learning outcome (J). To this end, as concerns H, we have used the data gathered in items 4, 5, 6 and 10. The results obtained are shown in Table 3 and are presented graphically in Figure 10. In relation to J, we have employed the final mark assigned to each task and the final mark obtained at the end of the activity.

Table 3. Pearson correlation coefficient ($n = 60$) for research question 2

<table>
<thead>
<tr>
<th></th>
<th>4.1.EA</th>
<th>4.2.EA</th>
<th>4.3.EA</th>
<th>4.4.EA</th>
<th>5.1.EA</th>
<th>5.2.EA</th>
<th>5.3.EA</th>
<th>5.4.EA</th>
<th>6.1.EA</th>
<th>6.2.EA</th>
<th>6.3.EA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</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>10.1.EA</td>
<td>.399**</td>
<td>-.070</td>
<td>-.212</td>
<td>-.329*</td>
<td>.300*</td>
<td>.143</td>
<td>.176</td>
<td>-.014</td>
<td>.492**</td>
<td>.174</td>
<td>-.156</td>
</tr>
<tr>
<td>10.2.EA</td>
<td>.235</td>
<td>-.078</td>
<td>.006</td>
<td>-.010</td>
<td>.299*</td>
<td>.005</td>
<td>.121</td>
<td>.074</td>
<td>.432**</td>
<td>.191</td>
<td>.139</td>
</tr>
<tr>
<td>10.3.EA</td>
<td>.213</td>
<td>-.082</td>
<td>.002</td>
<td>-.124</td>
<td>.310*</td>
<td>.213</td>
<td>.125</td>
<td>-.026</td>
<td>.275*</td>
<td>.180</td>
<td>-.108</td>
</tr>
<tr>
<td>10.4.EA</td>
<td>.537**</td>
<td>-.067</td>
<td>.316*</td>
<td>-.320*</td>
<td>.315*</td>
<td>.437**</td>
<td>.194</td>
<td>-.010</td>
<td>.338**</td>
<td>.466**</td>
<td>.207</td>
</tr>
<tr>
<td>10.5.EA</td>
<td>.236</td>
<td>-.199</td>
<td>-.164</td>
<td>.128</td>
<td>.278*</td>
<td>.271*</td>
<td>.071</td>
<td>.037</td>
<td>.471**</td>
<td>.266*</td>
<td>.055</td>
</tr>
<tr>
<td>EG</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>10.1.EA</td>
<td>.191</td>
<td>.046</td>
<td>-.124</td>
<td>-.045</td>
<td>.355**</td>
<td>.372**</td>
<td>-.013</td>
<td>-.143</td>
<td>.583*</td>
<td>.345**</td>
<td>-.193</td>
</tr>
<tr>
<td>10.2.EA</td>
<td>-.029</td>
<td>.077</td>
<td>-.052</td>
<td>.000</td>
<td>.325*</td>
<td>.478**</td>
<td>.045</td>
<td>.060</td>
<td>.453*</td>
<td>.194</td>
<td>-.290*</td>
</tr>
<tr>
<td>10.3.EA</td>
<td>.288*</td>
<td>.151</td>
<td>.074</td>
<td>-.076</td>
<td>.426**</td>
<td>.511**</td>
<td>.114</td>
<td>-.007</td>
<td>.290*</td>
<td>.225</td>
<td>-.257*</td>
</tr>
<tr>
<td>10.4.EA</td>
<td>.014</td>
<td>.069</td>
<td>.216</td>
<td>.098</td>
<td>.399*</td>
<td>.481**</td>
<td>.094</td>
<td>-.279*</td>
<td>.352**</td>
<td>.488**</td>
<td>-.271*</td>
</tr>
<tr>
<td>10.5.EA</td>
<td>.004</td>
<td>.173</td>
<td>.308*</td>
<td>.199</td>
<td>.295*</td>
<td>.457**</td>
<td>.042</td>
<td>.000</td>
<td>.496**</td>
<td>.288**</td>
<td>-.308*</td>
</tr>
</tbody>
</table>

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

As regards “students’ self-regulation,” EG students felt more motivated (86.33%) as well as more concentrated to the task (90.00%). Moreover, they showed more solidarity (85.00%) to their peers as well as more willingness to making suggestions (83.33%). All this allowed students to self-regulate their participation in the activity on time (91.67%) as well as to change to a more positive behavior faster (85.00%), a fact that allowed them to be
constructive and cooperative when they were facing socio-cognitive conflicts that occurred among the members of the group, due to the application of the cognitive dissonance strategy, and thus achieve the desired conceptual change more effectively. Furthermore, when students experienced joy (81.67%) they felt more involved to create and share knowledge on time (90.67%). As students were feeling more motivated and concentrated on the task, they felt less boredom (71.67%) or anger (68.33%), which led them to continuously try and thus improve their performance before it was too late (88.33%). Finally, high concentration combined with anxiety and opposed points of view seemed to influence positively a more balanced distribution of work among the group members, which achieved to lighten students’ workload (81.67%) during the development of the activity.

Students self-regulate their participation in the activity on time
Students change their behavior toward more positive faster
Students get involved to create and share knowledge on time
Students improve their performance before it's too late
Students lighten their workload

Figure 10. Correlation between students’ emotion awareness and self-regulation in (a) CG and (b) EG

CG students experienced less joy (71.67%), motivation (80.00%), concentration in the task (76.67%) and solidarity to peers (81.67%) than EG students. Under these circumstances, students managed to self-regulate their participation in the activity on time at a rate of 73.54%, whereas they had difficulty to show better ability to change to a more positive behavior faster (70.00%) as well as better skills to get involved to create and share knowledge on time (71.67%). Moreover, low performance in motivation, concentration and solidarity had negative effect in lightening students’ workload (75.00%), whereas when they felt emotions such as anxiety and anger, combined with low motivation, concentration, joy and solidarity, acted rather as a barrier to the students’ efforts to improve their performance before it was too late (81.67%).

As regards “learning outcome,” we were based on the data gathered in items 4, 5, and 6 of Figure 7 (X) and on the mark that students obtained in each task as well as on final mark awarded at the end of activity (J). The results obtained are shown in Tables 4, 5 and 6.

Table 4. Descriptive statistics of learning outcome

<table>
<thead>
<tr>
<th></th>
<th>CG (n = 60)</th>
<th></th>
<th>EG (n = 60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marks</td>
<td>Mean 7.5744</td>
<td>SD 1.47604</td>
<td>Skewness -.240</td>
</tr>
</tbody>
</table>

Table 4 shows that the average score of EG students has been higher than the one of CG students. In Table 5, when we examine all marks obtained by EG students, we see that there is a significant relationship between learning outcome and student’s supportive behavior with peers, which means that good learning outcome is achieved through solidarity among group members. Instead, examining all the marks obtained by CG students, we see that there is a significant relationship between the result obtained in the activity and a variety of emotions that students have experienced, ranging from joy to sadness or fear, and passing from concentration and solidarity with peers to opposition in case of disagreement. This oscillation in students’ emotions certainly influences their learning outcome in a positive or negative manner. Finally, our analysis obtained an interesting result that concerns the marks that were higher than or equal to 9 (Table 6). In the EG it was found again that there is a strong correlation between learning outcome and solidarity with peers. In the CG it was found that there is a strong correlation between learning outcome and opposition in case of disagreement. In this case, it is shown that strong oppositions among group members, without letting them have an explicit awareness of this situation, have prevented them from achieving a very good learning outcome.
Table 5. Pearson correlation coefficient and learning outcome with all the marks

<table>
<thead>
<tr>
<th></th>
<th>CG (n = 60)</th>
<th>EG (n = 60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All the marks</td>
<td>.328* .387** -.255* .546** .284* -.285*</td>
<td>.343* -.008 -.070 .581** .292* -.113</td>
</tr>
</tbody>
</table>

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

Table 6. Pearson correlation coefficient and learning outcome with marks greater than or equal to 9

<table>
<thead>
<tr>
<th></th>
<th>CG (n = 14)</th>
<th>EG (n = 27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marks &gt;= 9</td>
<td>-.616* .276*</td>
<td>.276*</td>
</tr>
</tbody>
</table>

Note. p = .05.

Regarding the third research question. We have correlated EA (X) with affective feedback (K). To this end, we have used the data gathered in items 4, 5, 6 and 8 of Figure 7. Table 7 shows the results obtained for items: 8.1.AF, 8.2.AF, 8.3.AF, 8.4.AF and 8.5.AF, whereas Figure 11 shows these results graphically. In fact, these items represent the competencies that teachers need to have in order to achieve a positive change on students’ affective and cognitive state.

Table 7. Pearson correlation coefficient for research question 3

<table>
<thead>
<tr>
<th></th>
<th>CG (n = 60)</th>
<th>EG (n = 60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1.AF</td>
<td>.444** .243 .409** .439** .270</td>
<td>.391** .285* .151 .202 .302</td>
</tr>
<tr>
<td>8.2.AF</td>
<td>-.235 .054 -.248 -.045 -.351**</td>
<td>-.008 -.001 -.013 .059 -.081</td>
</tr>
<tr>
<td>8.3.AF</td>
<td>-.186 .142 -.167 .367** .012</td>
<td>.296* .079 -.070 .024 -.229</td>
</tr>
<tr>
<td>8.4.AF</td>
<td>-.264* -.196 -.217 -.106 -.215</td>
<td>.192 .131 .028 .189 -.016</td>
</tr>
<tr>
<td>8.5.AF</td>
<td>.352** .409** .292* .635** .313*</td>
<td>.411** .469** .137 .320* .006</td>
</tr>
<tr>
<td>5.1.EA</td>
<td>.378** .232 .424** .269* .496**</td>
<td>.128 .288* .017 .214 .012</td>
</tr>
<tr>
<td>5.2.EA</td>
<td>.087 .148 -.078 .079 -.252</td>
<td>.099 .288* .038 .090 -.013</td>
</tr>
<tr>
<td>5.3.EA</td>
<td>.046 .038 -.118 .012 -.265*</td>
<td>-.208 .029 .106 .241 .225</td>
</tr>
<tr>
<td>5.4.EA</td>
<td>.559** .074 .402** .319* .408**</td>
<td>.161 .280* .311* .275* -.192</td>
</tr>
<tr>
<td>6.1.EA</td>
<td>.281* .311* .069 .469** .389**</td>
<td>-.179 -.083 .261* .123 .034</td>
</tr>
<tr>
<td>6.2.EA</td>
<td>-.031 .138 -.133 .459** .066</td>
<td>.010 .022 .319* -.003 -.007</td>
</tr>
<tr>
<td>6.3.EA</td>
<td>.001 .319* .066</td>
<td>.010 .022 .319* -.003 -.007</td>
</tr>
</tbody>
</table>

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

Figure 11. Correlation between teacher’s awareness about students’ emotions and his/her attitude and feedback in (a) CG and (b) EG

As regards item 8.1.AF, the teacher has extensively used dynamic methodologies to motivate students to learn (91.67%). As shown in the correlations in Table 7, this helped more CG students in several aspects (provided them more satisfaction, alleviated their frustration, supported more motivation, enhanced their concentration to the activity and solidarity with others as well as encouraged their involvement to the activity through suggestions and opinions).

As concerns item 8.2.AF, the teacher had to attend more the feelings and emotions of EG students, when there was a conflict in the group. Regarding 8.3.AF, the teacher has facilitated group discussion to manage emotions at 76.67%.
This action has helped EG students mainly in three aspects: increasing their solidarity, encouraging them to provide more suggestions and opinions, and mediating to resolve conflicts. Instead, CG students were mainly benefited in basic emotional aspects, since they felt happier, more motivated, concentrated and more sympathetic. Finally, exploring the other two items (8.4.AF and 8.5.AF), in which the teacher has encouraged and motivated students’ individual work sharing it with the team, as well as the teacher has resolved students’ questions offering advice and suggestions, these actions have helped much more CG students in many aspects as seen in Table 7. Regarding EG students, they do not seem to have very significant correlations (except from two aspects related to students’ motivation and solidarity which are encouraged by teacher’s action 8.4.AF).

Discussion

Given the three research questions we set in this work (for the sake of convenience we repeat them here), we analyze the results presented above and obtain the following conclusions:

(1) Is there a significant correlation between students’ emotion awareness and their motivation and engagement in learning?
(2) Is there any significant correlation between students’ emotion awareness and their self-regulation and learning outcome?
(3) Is there any significant correlation between teacher’s awareness about students’ emotions and his/her attitude and feedback?

In regard of the first research question, students in both groups showed high levels of motivation under the existence of positive emotions such as joy, as well as strong concentration to the task and solidarity to their peers. However, in the presence of not so positive emotions (such as sadness/shame, fear/anxiety, and anger/frustration), CG students felt very bored and high tendency to dispute, which led them to lose motivation to continue their activities. Moreover, they showed low solidarity to their peers. Finally, when they felt anger or frustration, they had a strong lack of self-confidence. In contrast, EG students when they felt negative emotions such as anxiety or frustration, little more than half of them felt unsafe, but they were able to maintain at least a minimum interest on the activity. This was even more obvious when they were feeling sad. Moreover, when these students felt sad, they were able to receive and provide suggestions and opinions in a constructive way, thus they managed to maintain their engagement during the development of the activity. As a consequence of the above, we draw the conclusion that there is a significant positive correlation between emotion awareness and students’ motivation and engagement in learning.

In regard of the second research question, CG students definitively obtain lower scores in self-regulating all explored aspects, that is, a more timely participation in the activity, the necessary changes that could lead towards a more positive behavior faster, a more timely involvement to create and share knowledge, a better performance before it's too late, and a more balanced distribution of their workload. In contrast, EG students achieved much better results in self-regulating all these aspects, distinguishing self-regulation skills such as timely participation and effective knowledge management that scored above 90% and which contributed to enhance teamwork and a more effective development of the activity. Considering learning outcome, EG students performed better than CG students. One of the reasons for achieving better learning outcome is grounded in building high degree of group solidarity and cohesion, which favours trust and engagement among the members of the group. Having the potential of emotion awareness of themselves and their peers during the whole activity provides students with an important tool to develop emotional competence for the group and thus build an emotionally intelligent team. As a result, we can claim that there is a significant positive correlation between emotion awareness and students’ self-regulation and learning outcome.

In regard of the third research question, the teacher had the capability to be aware of students’ emotions both in CG and EG. From the above results, we see that the teacher intervenes and supports both groups in almost all aspects that we explored. As regards the CG, since students in this group were not aware of their emotions, they needed much more support and affective feedback from their teacher, for this reason teachers’ attitude has been considered crucial in all aspects. Especially, teacher’s affective feedback has primarily involved dynamic methodologies to motivate students to learn, encouraged and motivated students’ individual work, sharing it with the team and resolved students’ questions offering advice and suggestions. At a second level, teacher has facilitated group discussion to manage emotions and attended students’ feelings and emotions when there was a conflict in the group. Regarding the EG, since the students were aware of their emotions all the time, they emphasized more their need to ask for
emotional support by the teacher when there was a conflict in the group. The result of the teacher intervention made these students feel happy, motivated, concentrated, safe, show more solidarity to their peers, encouraged to provide more suggestions and opinions, as well as more capable of resolving conflicts. All in all, this analysis proves that there is a significant positive correlation between emotion awareness and teacher’s attitude and feedback.

**Conclusions and future work**

In this work we investigated the way emotion awareness influences students’ motivation, engagement, self-regulation and learning outcome as well as teacher’s attitude and feedback (identifying, at the same time, the competencies that teachers need to have in order to achieve a positive change on students’ affective and cognitive state) in long-term blended learning practices. The results of our work showed that being aware of their emotions, students become more conscious of their situation, which prompts them to change and adapt their behavior for the benefit of their group. Moreover, it has been observed that their learning performance also improved in relation to their motivation, engagement and self-regulation. We also explored teachers’ attitude when they are conscious of students’ emotional state. We saw that teacher intervenes to support all students consciously and on time. In the case of Experimental Group students, teacher’s affective feedback becomes even more focused, knowing that students were aware of a difficult emotional situation they encountered in case of socio-cognitive conflict.

Our future work now turns to investigate more on the nature and impact of affective feedback on students’ learning process. This represents a first step toward the long-term objective of designing a virtual Affective Pedagogical Tutor (APT) that provides (semi)-automated feedback to students. This raises important issues about: the type of APT that is most appropriate and effective to student-centered collaborative learning situations; the understanding of the factors that have led students to remain in the same negative affective state for a certain period of time (which can lead them to deterioration of their learning performance, failure, and even withdrawal from studies); the establishment of criteria that indicates the most adequate moment that APT can intervene and monitor students’ affective situation; and, the type of affective feedback that best fits the students' needs and affective state.

**References**


The Effects of Positive and Negative Mood on Cognition and Motivation in Multimedia Learning Environment

Tze Wei Liew¹ and Su-Mae Tan²*

¹Department of Knowledge Management, Faculty of Business, Multimedia University, Malaysia // ²Centre for Diploma Programmes, Multimedia University, Malaysia // twliew@mmu.edu.my // smtan@mmu.edu.my

*Corresponding author

ABSTRACT
The Cognitive-Affective Theory of Learning with Media framework posits that the multimedia learning process is mediated by the learner’s mood. Recent studies have shown that positive mood has a facilitating effect on multimedia learning. Though literature has shown that negative mood encourages an individual to engage in a more systematic, elaborate, and analytical cognitive processing, this effect has not been empirically tested in the context of multimedia learning. The work is supported by two experiments. Experiment 1 was conducted to determine if the facilitating effects of positive mood found in previous works can be replicated with a multimedia learning system that teaches basic programming algorithm to learners at an Asian university. Experiment 2 investigated the effects of induced negative mood on learners’ cognition and motivation in a multimedia learning environment. Results revealed that negative mood enhanced intrinsic motivation and germane load, while reducing learning transfer. Learning transfer was enhanced by the presence of positive moods among males. Moreover, positive mood had a favorable effect on germane load, extraneous load, and intrinsic motivation. Theoretical and practical implications related to affective learning are discussed in this paper.

Keywords
Multimedia learning environment, Cognitive-Affective Theory of Learning with Media, Mood, Gender, Motivation

Introduction

In a typical e-learning and computer-based learning environment, learners are confronted with information presented in the format of text, narration, animation, and graphics. Hence, learners engage in the cognitive process known as multimedia learning (Mayer, 2008). The cognitive process of multimedia learning describes an information processing in which learners select relevant visual and verbal materials, organize these visual and verbal mental representations in coherent structures in working memory, and integrate the visual and the verbal mental representations with one another and with prior knowledge (Mayer, 2008).

Complementing this process model, the cognitive load theory (CLT) provides a capacity model of multimedia learning (Kalyuga, 2011). The CLT model highlights the inherent limitation of working memory and describes three types of cognitive load: intrinsic, germane, and extraneous. Intrinsic cognitive load is the inherent level of difficulty associated with a specific instructional topic. Germane load defines the amount of mental effort invested by the learner in comprehending the materials, whereas extraneous load describes processing demands of information that are not directly related to the learning task (Kalyuga, 2011).

Multimedia learning design has been primarily based on cognitive models such as dual-coding theory and cognitive load theory (Mayer & Moreno, 2003). However, Moreno’s (2006) Cognitive-Affective Theory of Learning with Media extended the framework by integrating emotional and motivational factors in multimedia learning design. According to this framework, the process of multimedia learning is mediated by emotional factors that influence learner’s willingness to invest mental effort to the learning process (Park et al., 2015). In addition, the framework assumes that multimedia learning process is mediated by learner’s metacognition such as cognitive styles and abilities (Moreno & Durán, 2004; Park et al., 2015). Hence, a learner’s working memory, motivation (i.e., germane load), and cognitive style is affected by his or her emotional states.

Recent studies have attempted to validate the importance of emotional factors in the design of multimedia learning environments. In a seminal work by Um et al. (2012), it was demonstrated that learners who were induced with positive mood outperformed learners who were induced with neutral mood in comprehension and transfer tests. The researchers also showed that positive emotion can be transferred to learners through interface designs, such as anthropomorphism and bright colors. In the attempt to replicate these findings, Plass et al. (2014) tested the effects of
color and anthropomorphism on learners’ emotion, motivational and cognitive outcomes. It was shown that positive emotional design in multimedia learning environments enhanced positive emotion, motivation, and comprehension. Moreover, positive emotional design in a multimedia interface significantly reduced learners’ perception of task difficulty. However, the study failed to replicate the facilitating effect of positive emotional design on transfer scores.

Mayer and Estrella (2014) tested the effects of positive emotional design in multimedia learning environments. Mirroring earlier findings (Plass et al., 2014; Um et al., 2012), results from their experiments showed that positive emotional design via anthropomorphism and appealing colors enhanced retention scores. In an eye tracking experiment, Park et al. (2015) revealed that learners who were induced with positive mood had higher learning outcomes in comprehension and transfer tests and showed longer fixation durations on the relevant information of the multimedia learning environment than learners who were induced with neutral mood. Their findings empirically indicate that emotional state influences cognitive and attentional processing in multimedia learning.

Moods in multimedia learning environment

There are different constructs when defining “emotion” (Ekkekakis, 2012; Russell, 2003). “Affect” is an umbrella term that bounds both “emotions” and “moods” (Russell, 2003). Moods differ from emotions in two key aspects. First, moods last longer than emotions (Ekkekakis, 2012). Secondly, moods are diffuse and operate at a lower intensity than emotions (Fridja, 2009). Unlike emotions, which are often prototypically labeled (e.g., surprise, fear, anger, and shame), moods can be generally described as feeling “good” or “bad” (Ekkekakis, 2012). While emotions are intense and trigger action-oriented responses, moods have a potentially more enduring, subtle, and insidious influence on people’s cognitive processes than do distinct and intense emotions (Forgas, 1995).

In addition, Brave and Nass (2003) provide a research framework for distinguishing the effects of emotion and mood in human-computer interaction. The researchers described the differences between emotion and mood from a functional perspective. Emotions bias action as they prepare the body and mind for an apt and instantaneous response, whereas mood affects an individual’s sentiment and cognitive style over a longer duration. Taking this functional viewpoint, we focused on mood rather than emotion in this research for three reasons.

 Firstly, interacting in multimedia learning environments often requires a substantial amount of time. Hence, we focused on the effect of moods experienced by learners as enduring states (e.g., feeling “good” versus “bad”) rather than short-lived emotions (e.g., surprise, fear, anger). Secondly, the experience of mood is constantly influenced by physical surroundings (Forgas, 2013). A recent framework by Choi et al. (2014) posits that environmental factors, such as lighting, noise, and temperature, influence learners’ mood when studying in multimedia learning environments. Hence, a learner’s mood (rather than emotion) is regarded as a subtle and enduring experience that is caused by environmental attributes. Thirdly, our focus on mood helps to link potential findings to the design of multimedia learning environments. As shown in literature, interface design, such as colors, website layouts, decorative graphics, and anthropomorphism, affect learners’ emotional and arousal states. Because these design factors are meant to be subtle and not overly seductive or distracting (see Magner et al., 2014; Mayer & Estrella, 2014), interface design factors tend to bias learners’ mood more than intense emotion.

The Cognitive-Affective Theory of Learning with Media provides us with the insights of how mood might mediate cognition and motivation in multimedia learning environment (Moreno, 2006; Park et al., 2015). From this framework, three crucial cognitive and metacognitive aspects are highlighted: working memory, motivation (i.e., germane load), and processing style. The following sections provide theoretical reviews related to mood, working memory, motivation, and processing style.

Effects of mood on working memory

Literature has predominantly shown that negative mood impairs cognitive performance by depleting resources in working memory (Brose et al., 2012). One argument is that negative mood prompts an individual’s mind to entertain irrelevant and ruminative thoughts (Nolen-Hoeksema et al., 2008). Moreover, sad individuals will engage in mood-repair strategy to alleviate negative mood, resulting in split attention in working memory between cognitive task and mood correction (Brand et al., 2007; Riediger et al., 2011). Lastly, the resource allocation hypothesis (Ellis et al.,
1988) argues that sad individuals tend to lessen their cognitive effort for difficult tasks that potentially further erode their mood.

However, research has also indicated that positive mood may interfere with working memory (see Mitchell & Phillips, 2007). The activation of positive mood has been shown to impair memory tasks, such as Tower of London (Phillips et al., 2002), the complex Stroop task (Phillips et al., 2002), and the running span task (Cowan et al., 2005). Drawing parallels to negative mood, positive mood may also prime individuals to entertain irrelevant thoughts, causing reduction in working memory capacity (Seibert & Ellis, 1991). This notion may explain the association between positive mood and increased distractibility during cognitive tasks (see Biss & Hasher, 2011).

**Effects of mood on motivation**

Affect-as-information theory (Schwarz & Clore, 1996) suggests that happy individuals are more likely to use heuristics to interpret their mood in order to evaluate their external environment (see Forgas, 2013). Hence, positive mood primes individuals to elicit favorable judgments regarding the target stimuli, resulting in enhanced motivation and engagement towards tasks (Isen & Reeve, 2005). In the context of multimedia learning environments, positive mood has been associated with increased intrinsic motivation and germane load in learners (Plass et al., 2014; Um et al., 2012).

Negative mood has been shown to increase effort and motivation in cognitive tasks (Erber & Erber, 1994; Forgas, 2013). The hedonistic discounting hypothesis argues that sad mood increases an individual’s expectation of future hedonistic value of a task, hence prompting him/her to invest more effort in the task (see Forgas, 2013). This notion may be related closely with control theory (Carver & Scheier, 1990); that is, negative affect may signal to individuals that their current progress is under certain expectations, thereby encouraging them to invest higher levels of motivated effort to reach their goals. Moreover, as part of mood-repair strategy, sad individuals tend to recruit effortful and motivated processing in tasks, in order to provide distraction from negative thoughts (Erber & Erber, 1994; Erber & Tesser, 1992; Forgas, 1995; Forgas, 2013).

**Effects of mood on processing style**

In his book, The Emotion Machine, Marvin Minsky (2006) states that when we change our emotional states, we are switching between different ways to think. From the literature, there is a general consensus suggesting that positive mood leads an individual to use less effortful and heuristic information processing, while negative mood recruits a more systematic, analytical, and elaborate style of thinking (see Forgas, 1995; Forgas, 2013). Forgas (2013) asserted that “negative mood, by recruiting more accommodative and externally focused processing, should improve attention and encoding” (p. 226). Hence, it has been contended that negative mood may benefit learning through the recruitment of more effortful thinking (Forgas, 1995; Forgas, 2013), elaborate memory encoding (Bäuml & Kuhbandner, 2007; Forgas et al., 2009) and cautious interpretation of new information (Bless & Fiedler, 1995; Gasper, 2003).

However, it has also been highlighted that negative mood causes an individual’s attention focus to be too narrow and rigid when processing external data, thus interfering with flexible thinking (Gasper, 2003). On the contrary, the build and broaden hypothesis (Fredrickson & Branigan, 2005) suggests that attention span and global processing of external information improves in the presence of positive mood. Under this view, happy individuals are more likely to exhibit flexible thinking that allows for more global integration and creative analysis of newly acquired information (Gasper, 2003), thus resulting in enhanced creativity and problem solving skills (Isen, 1999).

**Gender as moderating factor to effects of mood**

Negative and positive moods are processed differently by males and females (Martin, 2003). Martin (2003) notes that “females have a greater propensity for detailed processing, whereas males tend toward more heuristic processing” (p. 253). As a consequence, males and females approach mood-repair differently. Sad males engage in mood-repair by
using distraction and therefore are less likely to pay attention to difficult tasks that potentially worsen their mood (Nolen-Hoeksema, 2001). On the other hand, under the influence of negative mood, females tend to recruit detailed, comprehensive, and elaborate thinking in processing external environments as a whole, including stimuli that potentially worsen mood (Martin, 2003). This is to say that sad females are more likely to engage in more motivated and effortful processing when dealing with difficult tasks compared to sad males.

Research gaps and research objectives

It is tempting to intuitively proclaim that positive mood will always enhance cognitive performance and motivation in the context of learning, while negative mood will hinder cognition and motivation. However, as discussed in the previous section of this paper, the findings from psychological research showed that the association among mood, cognition, and motivation remains contradictory. For instance, though initial studies has shown the benefits of positive mood with regards to multimedia learning (Park et al., 2015; Plass et al., 2014; Um et al., 2012), literature from cognitive psychology has also pointed out the association among positive mood, decreased working memory and increased distractibility during cognitive tasks (Biss & Hasher, 2011; Mitchell & Phillips, 2007).

As noted by Heidig et al. (2015), research related to the effects of mood on multimedia learning is still in its infancy. Thus far, the facilitating effect of positive mood has been demonstrated for a representative sample of students from Western (Mayer & Estrella, 2014; Um et al., 2012) and European universities (Park et al., 2015; Plass et al., 2014). However, the effects of positive mood on multimedia learning have not been tested for learners in Asia. Moreover, the same multimedia learning topic (i.e., immunization) was used in previous studies (Mayer & Estrella, 2014; Park et al., 2015; Plass et al., 2014; Um et al., 2012). As highlighted by Mayer and Estrella (2014), there is, therefore, a need to extend this research by using different subject domains and sample populations. Hence, the first objective of this study is to determine if the facilitating effects of positive mood found in previous works can be replicated in a multimedia learning system that teaches basic programming algorithm to learners from an Asian university.

Secondly, the role of negative mood in multimedia learning environments is still unclear (Heidig et al., 2015). While negative mood is generally undesirable, its potential to promote a more systematic, elaborate, and analytical cognitive processing (Bless & Fiedler, 1995; Forgas, 2013; Gasper, 2003) should be tested in the context of multimedia learning. Furthermore, negative mood may act as a signal for an individual to put in more effort (Erber & Erber, 1994) and encourage learners to be cautious and pay more attention to information in the environment (Bless & Fiedler, 1995; Forgas, 2013; Gasper, 2003). As there are currently no empirical studies that investigate the effects of negative mood on multimedia learning, it is uncertain whether negative mood can enhance or decrease learners’ motivation and cognition in multimedia learning environments. Hence, the second objective of this study is to determine the effects of negative mood on learners’ cognition and motivation in a multimedia learning environment.

Previous mood researches tend to directly compare between positive and negative mood, rather than comparing them with neutral mood (control group). However, because positive and negative mood has been shown to trigger different systems of cognition and motivation in individuals, control conditions (i.e., neutral mood) are crucial to isolate the effects of positive and negative mood. For example, if both positive mood and negative mood are assumed to enhance motivation, a direct comparison between these two conditions (i.e., absence of a neutral condition) may yield obscure findings. Hence, we included the control conditions in two separate experiments: (1) positive mood versus neutral mood and (2) negative mood versus neutral mood, in order to isolate the effects of positive and negative mood, respectively. Additionally, both experiments were conducted separately because we used different surveys for measuring positive mood and negative mood in learners.

Research questions

From the literature, mood theories offer contradictory predictions regarding the effects of mood on cognitive load, germane load, intrinsic motivation, and learning performance in multimedia learning systems. Hence, we aim to answer the following questions:

(1) How does positive mood affect cognitive load, germane load, intrinsic motivation, and learning performance in multimedia learning environments?
Does gender differently affect the impact of positive mood on cognitive load, germane load, intrinsic motivation, and learning performance in multimedia learning environments?

(3) How does negative mood affect cognitive load, germane load, intrinsic motivation, and learning performance in multimedia learning environments?

(4) Does gender differently affect the impact of negative mood on cognitive load, germane load, intrinsic motivation, and learning performance in multimedia learning environments?

**Experiment 1**

We aimed to compare the effects of positive mood and neutral mood on learner’s cognition and motivation in multimedia learning environments.

**Participants**

Freshmen ($n = 172$: 88 males, 84 females; mean age = 19) at a university were recruited to participate in exchange for extra course credits. All participants were business majors who had no experience in programming-related subjects.

**Multimedia learning environment**

The multimedia learning environment delivered lessons on basic computer programming (see Figure1). Participants were required to learn to (1) comprehend computer algorithms of if-statements, if-else statements, and nested-if statements and (2) predict the output from C-Programming statements. Within the multimedia learning environment, participants first needed to study the fundamental program concepts, such as initial value, condition, and statement. To be able to predict program outputs, participants must form correct mental models on program algorithms by grasping the relationship between the initial value, condition, and statement. Sample programs and flowcharts were included to highlight these relationships, and subsequent working demonstrations were presented to show how to predict the program output. The ability to identify programming concepts and to predict program outputs are consistent with the cognitive skills of knowledge retention and comprehension of Bloom’s taxonomy in computer sciences (as cited in Thompson et al., 2008).

![Figure 1. Multimedia learning environment](image)

**Dependent measures**

*Positive mood*

To measure a learner’s positive mood, the positive affect scale (PAS) from PANAS-X was utilized (Watson et al., 1988). This survey has been formerly used in studies that examine emotional responses to interface design in
multimedia learning environments (Plass et al., 2014; Um et al., 2012). The survey asked respondents to indicate the degree to which they experienced 10 types of positive feelings, using a 7-point Likert-type scale ranging from 1 (very slightly or not at all) to 7 (very much; coefficient alpha = .84). The total score for positive affect was obtained by adding the 10 responses from the items.

**Intrinsic motivation**

The learner’s motivation was measured using a self-report instrument, consisting of an 8-item questionnaire with 7-point Likert style items, developed by Isen and Reeve (2005) for measuring intrinsic motivation. Participants were asked to rate how interesting and enjoyable they found the experience (1 = strongly disagree, 7 = strongly agree). One point was assigned for each item, and each participant’s total score, obtained by adding responses from the 8 items, was used for the data analysis. The measure was used by prior studies related to multimedia learning (Plass et al., 2014; Um et al., 2012).

**Learning performance**

The learning performance of participants was measured as a total score of the transfer post-tests. The comprehension test asked learners to predict the output of ten short program statements. The transfer tests require learners to apply algorithmic rules as learned in the virtual lessons to the new situation, in order to achieve the correct output. The maximum score obtainable is 10. These measures were consistent with the cognitive skills of knowledge retention and comprehension in computer sciences (Thompson et al., 2008).

**Learners’ germane and extraneous load**

To measure the cognitive load experienced by learners, participants completed a 9-point Likert style Cognitive Load Subjective Experience Questionnaire targeting invested mental effort (Paas, 1992) and a 7-point Likert style survey on their perceptions of task difficulty (Kalyuga et al., 2000).

**Procedure**

Participants were randomly divided and ushered to into either one of two identically designed lecture rooms, thereby forming the control group or the experimental group. To induce neutral and positive mood, the Velten procedure was administered (Velten, 1968; Jennings et al., 2000). Participants in the control group were asked to read 30 positively valenced statements twice, while participants in experimental group were asked to read 30 neutrally valenced statements twice (i.e., statements which were neither self-referent nor pertaining to mood). After the mood inductions, the PAS survey was administered. Next, participants viewed and listened to a 20-minute virtual lesson that was projected onto a 7000 mm x 7000 mm screen with attached PA speakers. After the virtual lesson, participants were instructed to complete the surveys for measuring positive affect, intrinsic motivation, cognitive load, and germane load. They were then given 20 minutes to complete the transfer posttest. Using a scoring rubric, each learner’s posttest was scored blind with respect to experimental conditions set by the second author of this study.

**Experiment 2**

We compared the effects of between negative and neutral mood on cognition and motivation in the multimedia learning environment. Freshmen (n = 96: 40 males, 56 females; mean age = 19) who had no programming prior knowledge were recruited to participate in exchange for extra course credits. The details and the procedure of this experiment are identical to Experiment 1 except for two variations. First, the negative affect scale (NAS) instead of the positive affect scale (PAS) from PANAS-X was used. Secondly, to induce a state of negative mood, 30 negatively valence statements were read by participants twice in the experimental group (Jennings et al., 2000).
Research question revisited

RQ1: How does positive mood affect cognitive load, germane load, intrinsic motivation and learning performance in multimedia learning environment?

Based on data from Experiment 1, we conducted a series of independent samples t-tests to compare scores for PAS 1 (administered immediately after mood induction), PAS 2 (administered after virtual lesson), perceived task difficulty, mental effort, and posttest scores between learners in the control (neutral mood) and experimental (positive mood) groups. Table 1 shows the means, standard deviations and p-values of the t-test analysis.

<table>
<thead>
<tr>
<th></th>
<th>Control (Neutral Mood) [Mean, SD]</th>
<th>Experimental (Positive Mood) [Mean, SD]</th>
<th>p-value (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAS 1 (after mood induction)</td>
<td>34.50 (9.77)</td>
<td>41.69 (9.78)</td>
<td>.000</td>
</tr>
<tr>
<td>PAS 2 (after virtual lesson)</td>
<td>34.76 (11.49)</td>
<td>41.13 (12.05)</td>
<td>.000</td>
</tr>
<tr>
<td>Perceived task difficulty (Extraneous load)</td>
<td>3.61 (1.51)</td>
<td>3.12 (1.53)</td>
<td>.061</td>
</tr>
<tr>
<td>Mental effort (Germane load)</td>
<td>3.71 (1.38)</td>
<td>4.62 (.91)</td>
<td>.000</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>29.05 (8.88)</td>
<td>34.23 (9.95)</td>
<td>.000</td>
</tr>
<tr>
<td>Posttest scores</td>
<td>5.78 (3.39)</td>
<td>6.05 (3.15)</td>
<td>.582</td>
</tr>
</tbody>
</table>

Positive mood induction was shown to be successful; PAS 1 scores for learners in the experimental group were significantly higher than PAS 1 scores for the control learners. Pairwise comparisons using paired samples t-tests showed no significant differences between PAS 1 and PAS 2 for the control learners, \( t(75) = 2.411, p = .759 \) and learners in the experimental group, \( t(95) = 2.411, p = .547 \). That is, for both the control and experimental groups, the levels of positive mood in participants were not affected by the learning task in the multimedia learning environment.

Positive-mood learners reported significantly higher mental effort, \( t(170) = 2.411, p = .000 \), and intrinsic motivation, \( t(170) = 2.411, p = .000 \) than did negative-mood learners. For perceived task difficulty, participants who received positive mood induction rated the difficulty of the learning material as lower than learners who received neutral mood induction, \( t(170) = 2.411, p = .061 \). Posttest scores did not significantly differ between learners in experimental and control groups.

RQ2: Does gender differently affect the impact of positive mood on cognitive load, germane load, intrinsic motivation, and learning performance in multimedia learning environments?

Using the data from Experiment 1, we conducted a multivariate analysis of variance (MANOVA) with mood condition (neutral, positive) and gender as between-subjects factors and perceived task difficulty, mental effort, intrinsic motivation, and posttest scores as dependent measures. The analysis showed there was an interaction effect between gender and mood condition for posttest scores, \( F(1,168)= 4.104, p = .044 \), partial eta squared = .024, and for mental effort, \( F(1,168) = 12.644, p = .008 \), partial eta squared = .041.

In terms of posttest scores, pairwise comparisons using t-tests revealed that positive-mood males (\( M = 7.43, SD = 2.08 \)) significantly outperformed neutral-mood males (\( M = 6.09, SD=3.41 \)), \( t(66.68) = 2.194, p = .032 \) and positive-mood females (\( M = 4.78, SD = 3.43 \)), \( t(81.94) = 4.619, p = .000 \). For mental effort, post-hoc analysis revealed that positive-mood females (\( M = 4.84, SD = 1.39 \)) significantly invested higher mental effort than did neutral-mood females (\( M = 3.35, SD = 1.20 \)), \( t(82) = 5.075, p = .000 \). Taken together, it was found that positive mood enhanced learning performance in male learners, while promoting a recruitment of higher mental effort in female learners.

RQ3: How does negative mood affect cognitive load, germane load, intrinsic motivation, and learning performance in multimedia learning environments?

Independent samples t-tests were conducted using the data from Experiment 2 to compare scores for NAS 1, NAS 2, perceived task difficulty, mental effort, and posttest scores between learners in the control (neutral mood) and
experimental (negative mood) groups. Table 2 shows the means, standard deviations, and p-values of the t-test analysis.

<table>
<thead>
<tr>
<th></th>
<th>Control (Neutral Mood)</th>
<th>Experimental (Negative Mood)</th>
<th>p-value (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAS 1 (after mood induction)</td>
<td>21.41 (5.29)</td>
<td>44.41 (9.18)</td>
<td>.000</td>
</tr>
<tr>
<td>NAS 2 (after virtual lesson)</td>
<td>17.50 (9.42)</td>
<td>26.91 (11.95)</td>
<td>.000</td>
</tr>
<tr>
<td>Perceived task difficulty (Extraneous load)</td>
<td>2.83 (1.56)</td>
<td>3.37 (1.74)</td>
<td>.113</td>
</tr>
<tr>
<td>Mental effort (Germaine load)</td>
<td>3.62 (1.29)</td>
<td>4.25 (1.78)</td>
<td>.052</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>23.25 (9.32)</td>
<td>29.25 (9.47)</td>
<td>.002</td>
</tr>
<tr>
<td>Posttest scores</td>
<td>6.42 (3.91)</td>
<td>3.92 (3.21)</td>
<td>.000</td>
</tr>
</tbody>
</table>

Like in Experiment 1, the negative mood induction was proven to be successful; learners who received negative mood induction reported higher negative affect scores (NAS 1) than did learners who received neutral mood induction. $t(75.15) = 15.02, p = .000$. Paired-samples t-tests revealed that the levels of negative affects dropped significantly from NAS 1 to NAS 2, for both control groups, $t(47) = 2.75, p = .008$ and experimental groups, $t(47) = 8.65, p = .000$. This suggested that the learning process in the virtual learning system activated mood repair strategies that lessen negative mood in learners.

With respect to posttest scores, negative-mood learners performed significantly worse than neutral-mood learners, $t(94) = 3.94, p = .000$. The magnitude of the differences in the means was large (eta squared = .14). Hence, the notion that negative mood can improve learning by recruiting a more systematic, analytical, and motivated processing was not supported by our data. However, we observed that negative-mood learners recruited higher mental effort, $t(94) = 1.96, p = .052$ and intrinsic motivation, $t(94) = 3.13, p = .002$ than did neutral-mood learners. This finding supported the notion that negative mood increases motivation and mental efforts in cognitive tasks.

**RQ4: Does gender differently affect the impact of negative mood on cognitive load, germane load, intrinsic motivation and learning performance in multimedia learning environments?**

A 2(gender) x 2(mood condition) MANOVA was performed, with perceived task difficulty, mental effort, intrinsic motivation and posttest scores as dependent measures. There was a significant interaction between gender and mood conditions for intrinsic motivation, $F(1.90) = 25.59, p = .000$, partial eta squared = .221. Following the guidelines by Cohen (2013), the value of the partial eta squared suggests a large effect. Post-hoc comparisons revealed that negative-mood females ($M = 34.35, SD = 7.04$) reported higher intrinsic motivation than did neutral-mood females ($M = 20.53, SD = 7.82$), $t(52) = 6.83, p = .000$. Whereas, there was no significant difference between the level of intrinsic motivation in negative-mood males ($M = 23, SD = 7.61$) and neutral-mood males ($M = 26.45, SD = 10.08$), $t(38) = 1.199, p = .238$.

Moreover, it was also found that in the neutral-mood condition, males ($M = 26.45, SD = 10.08$) reported higher intrinsic motivation than did females ($M = 20.53, SD = 7.82$), $t(46) = 2.287, p = .027$. Whereas in the negative-mood condition, females ($M = 34.35, SD = 7.04$) attributed higher intrinsic motivation than did males ($M = 23, SD = 7.61$), $t(44) = 5.174, p = .000$. As part of exploratory analysis, planned comparison t-tests (split by gender) revealed that the presence of negative mood significantly enhanced mental efforts in females (negative mood, $M = 4.35, SD = .91$; neutral mood, $M = 3.53, SD = 1.3$), $t(44.387) = 2.656, p = .01$. With respect to mental effort, no significant difference was found between negative-mood males ($M = 3.77, SD = 2.51$) and neutral-mood males ($M = 3.72, SD = 1.31$), $t(24.5) = .077, p = .935$. Taken together, it was shown under the influence of negative mood; female learners were more likely to increase their intrinsic motivation and mental effort in learning task than did male learners.
Discussion

Consistent with the findings from previous related studies (Mayer & Estrella, 2014; Park et al., 2015; Plass et al., 2014; Um et al., 2012), our data from Experiment 1 demonstrated the facilitating effect of positive mood on learner’s cognition and motivation in multimedia learning environment. Specifically, we observed that the induction of positive mood enhanced learner’s intrinsic motivation and mental effort in the multimedia learning system. It was also shown that the presence of positive mood lowered the perception of task difficulty in learners. This finding supports the affect-as-information theory (Schwarz & Clore, 1996). That is, learners who were put under the influence of a positive mood employed their interpretation of their current mood to guide their judgment, thus giving rise to positive evaluations regarding the multimedia lesson itself. As a result, happy participants were more likely to recruit higher mental effort and intrinsic motivation when interacting with the multimedia lesson than did the control participants.

Moreover, positive mood was found to improve learning transfer, albeit for male learners only. Hence, the notion that positive mood may stifle working memory by causing distraction (Biss & Hasher, 2011; Mitchell & Phillips, 2007; Riediger et al., 2011) was not supported in this study. Rather, the results supported the hypothesis that positive mood may enhance learning, plausibly through the enhancement of learners’ attention span, thought flexibility, and global integration of external information (Fredrickson & Branigan, 2005; Gasper, 2003). These heightened cognitive aspects may be particularly useful when interacting with multimedia learning systems, as multimedia learning involves attention and integration of multiple sources of information, i.e., narration, animation, and pictures (Mayer, 2008).

On the other hand, learning transfer was impaired under the influence of negative mood (Experiment 2). Learners who were induced with negative mood performed significantly worse in transfer test than did control learners. Therefore, our data generally did not fit with the notion that negative mood can promote learning by encouraging a more systematic, analytical, and comprehensive cognitive processing to new information (Forgas, 2013). In the context of our experiment, it is unclear whether sad learners might have been overly focused on processing certain data from the multimedia learning environment, causing thinking process to be too narrow and rigid to effectively integrate and relate new information in novel ways (see Gasper, 2003).

Moreover, the presence of negative mood might have caused the cognitive aspect of learners to account for both learning process and mood-repair, resulting in a decrease in working memory capacity (Brand et al., 2007; Riediger et al., 2011). The fact that learners in Experiment 2 experienced a significant decline of negative affect after the multimedia lesson provided evidence of mood-repair during multimedia learning. Nevertheless, it is interesting to note that negative mood heightened learners’ intrinsic motivation and willingness to invest mental effort in multimedia learning. This effect was more pronounced among females than males, plausibly because females are more likely to engage in detailed processing while males tend to limit their cognitive effort, thus relying on heuristic processing (see Martin, 2003).

Why did negative mood encourage learners to invest higher mental effort in multimedia learning? Two reasons may account for this observation. Firstly, negative mood could have led learners to recruit a more motivated and effortful processing as part of their mood repair mechanism (Erber & Erber, 1994; Forgas, 2013). Because the cognitive process in multimedia lesson is potentially engaging, sad learners took the opportunity to draw their mental focus on learning, while leaving little mental space for negative rumination. Secondly, according to hedonistic discounting theory (see Forgas, 2013), negative mood might have increased learners’ expectation of hedonistic value of future success, which resulted in heightened perseverance and intrinsic motivation towards the cognitive task in the multimedia environment. Perhaps, sad learners might have used their negative mood as information to interpret their status as underperforming, which prompted them to invest higher cognitive endeavor towards the multimedia lesson to achieve a sense of self-worth and achievement.

As discussed earlier, we like to point out that negative mood may be beneficial in cautioning learners to recruit more effort and deliberate processing in cognitive tasks. Forgas (2013) reported that sad participants demonstrated higher persistence and answered more questions in a challenging cognitive task than happy participants. Furthermore, as negative mood signals that the environment is problematic, Gasper (2003) notes that a more cautious approach is used by learners when paying attention to new to update and alter their knowledge base. Hence, if we assume that negative mood causes a learner to put in extra effort and vigilance in cognitive task, can multimedia learning benefit
from negative mood? Effectively, our experiment design involved a 20-minutes multimedia presentation that did not allow learners to control the pace or repeat the multimedia instructions. Therefore, if learners were given the flexibility to control the pace and study the multimedia material for as long as they like, we wonder if negative mood would have encouraged learners to spend more time and repetition in the multimedia learning environment, resulting in enhanced transfer performance. Further studies should be conducted to seek clarification regarding this issue.

Conclusion

Our study generally showed that positive mood had a facilitation effect on cognition and motivation in a multimedia learning system, whereas negative mood had a detrimental effect on multimedia learning. These findings demonstrated that positive mood is desirable for learning in multimedia learning systems, while the presence of negative mood should be mitigated. How can multimedia learning be designed to promote positive mood and to reduce negative mood? One approach is to focus on the visceral design of the multimedia learning interface to induce positive mood. There is a recent research trend that posits emotional design and aesthetic factors, such as colors, anthropomorphism, layouts, and illustrations in multimedia learning systems (Heidig, 2015; Plass et al., 2014; Um et al., 2012).

Yet another approach is to design emotionally intelligent systems that interpret learners’ mood profile and provides appropriate feedback as part of a mood regulation strategy (see Sottilare & Proctor, 2012). There is also a huge potential for embodied emotional agents to provide social and emotional support necessary for mood corrections in multimedia learning systems. To conclude, we highlight the need for future works to further explore the relationship among mood, cognition, motivation, and emotion-adaptive strategies in multimedia learning environment. For instance, because negative moods are thought to recruit systematic, cautious and deliberate processing patterns when analyzing new data (Gasper, 2003), it would be interesting to determine if negative moods may actually promote cognitive performance in simulation-based systems that require experimentation, hypotheses testing and discovery-based learning (Liew et al., 2014). In addition, research on how virtual agents and interfaces can be designed to regulate learners’ moods in learning systems warrants further attention.

Acknowledgements

This research has been supported by the Fundamental Research Grant Scheme by Malaysia Ministry of Higher Education under grant agreement ID: MMU/RMC-PL/SL/FRGS/2014/1.

References


Designing Empathetic Animated Agents for a B-Learning Training Environment within the Electrical Domain

Yasmín Hernández¹, Miguel Pérez-Ramírez¹, Ramón Zatarain-Cabada², Lucia Barrón-Estrada² and Giner Alor-Hernández³

¹Instituto de Investigaciones Eléctricas, Gerencia de Tecnologías de la Información, Cuernavaca, México // ²Instituto Tecnológico de Culiacán, Departamento de Posgrado, Culiacán, México // ³Instituto Tecnológico de Orizaba, División de Estudios de Posgrado e Investigación, Orizaba, México // myhp@iie.org.mx // mperez@iie.org.mx // rzatarain@itculiacan.edu.mx // lbarron@itculiacan.edu.mx // galor@itorizaba.edu.mx

*Corresponding author

ABSTRACT

Electrical tests involve high risk; therefore utility companies require highly qualified electricians and efficient training. Recently, training for electrical tests has been supported by virtual reality systems; nonetheless, these training systems are not yet adaptive. We propose a b-learning model to support adaptive and distance training. The adaptation is based on a representation of the knowledge and affect of the trainee. One of the main components is an animated pedagogical agent who guides trainees and provides instruction. The agent deploys different facial expressions conveying emotion and empathy to trainees. We incorporated these constructs into the learning model aiming to achieve credibility, social interaction, and user engagement, and thus improve learning. To develop the agent, we included some of the basic characteristics of electricians: work uniform and safety gear, such as helmets, gloves, and glasses, among other features. We believe that by representing the tutor as an electrician, the training environment will be better accepted by operators. The focus of this paper is a study that we conducted to gather knowledge about the expressivity and appearance of the animated agent. The participants evaluated the facial expressions of the agent and stated their point of view regarding the incorporation of the animated agent into the training system. The gathered knowledge is being used to enhance the animated agent before completing its integration to the training environment.

Keywords

Electrical tests, Virtual reality, Blended learning, Affective student model, Empathetic agents

Introduction

Performing electrical tests involves dangerous activities. Therefore, efficient training is needed in order to minimize training time, costs, and equipment damage, and most importantly, to prevent accidents that could injure electricians. However, training on electrical tests faces some problematic situations, such as limited opportunity to practice in an actual substation, since substations are operating on a regular basis.

We have developed a Virtual Reality System (VRS) to support traditional training. Trainees still attend classroom courses but they complement their learning and practice aided by the VRS. The system is used at the Mexican electricity utility by a number of electricians. However, the VRS is not adaptive to trainees’ knowledge and emotional states. Training still depends on a courses schedule, which most of the times, means waiting for a training course for a long time.

Based on this VRS that supports traditional training, we composed a blended learning model. Blended learning is a new strand in education, although the concept itself has already existed for a long time. It can take many forms and there are several definitions that include the same elements roughly. In a formal education context, blended learning is a formal educational program in which a student learns, at least in part through online delivery of content and instruction with some elements of student control over time, place, path, or pace, and at least in part at a supervised brick-and-mortar location away from home (Staker & Horn, 2012). A more general and perhaps more accepted definition states that blended learning is learning that is facilitated by the effective combination of different delivery modes, teaching models and learning styles, and founded on transparent communication amongst all parties involved in a course (Heinze & Procter, 2004).

In addition to blended learning, the characteristics of electrical tests demand faster and more efficient training, where instruction is adaptive and responds to needs of the trainees. For that reason, we have defined a roadmap to achieve
efficient and tailored training. The roadmap starts with traditional training and ends with intelligent and adaptive training.

In order to have more adaptive training, we have built a trainee model. This model represents the affective and knowledge states of trainees and allows the training system to adapt the instruction to a particular trainee. Our proposal for the affective state is based on facial expressions and contextual information.

We propose to present the instruction by an animated agent representing the instructor. We are designing an empathetic agent to try to engage the trainee, promote a positive emotional state and improve learning. A key highlight of the agent’s design is that it is presented as an electrician so that trainees can feel identified with him.

With all these elements, we are developing a model and a platform for blended training, which adapts itself to needs of the trainee in an intelligent way based on the roadmap for training. In this paper we present the outcomes of a study that we conducted on animated agents in order to refine our empathetic agent before full integration to the VRS. Experts in electrical tests who know the VRS participated in the study. We gathered useful information to improve the design of the agent, which includes ingredients such as appearance, behavior, and facial expressions.

The rest of the paper is organized as follows: the next section presents some related work. Then, we present the roadmap for training, describing the virtual reality system and the elements of the blended training model, particularly, the affective student model and the empathetic agent. After that, we present the evaluation of the empathetic agent. Then, results are analysed. Finally, we present some conclusions, as well as a description of future work.

**Related work**

Nowadays, training is a strategic activity in corporations as it is recognized that the efficiency of organizations depends directly on human capital, which in turn may depend on adequate training. High productivity is the result of efficient training, which becomes even more valuable when there is risk of accidents that harm people. Such is the case of electrical tests, which involve risk of electric shock, arc flash and other hazards; also there may be potential damage to the primary equipment in the substation. Several technologies have been used to improve training.

**Virtual reality and training**

Virtual Reality (VR) has demonstrated to be successful for training in several fields, among others, medicine (Dev & Heinrichs, 2008) and military applications (Reger, Rizzo & Gahm, 2015) where hazard and possibility of damage are involved. For instance, John (2008) presents a virtual environment designed to train in the Seldinger Technique, a common procedure in interventional radiology. Karadogan and Williams (2013) report the development of a tool based on haptics and virtual reality technology to augment the teaching of palpatory.

In education and training, VR allows the construction and integration of different learning contexts that make it successful as a learning tool (Pérez-Ramírez & Ontiveros Hernández, 2009, Pantelidis, 2009). In accordance with pioneer works (Burdea & Coiffet, 2003) and recent works (Fuchs, Moreau & Guitton, 2011), training is one of the main fields for VR application, as it possesses characteristics that are either inexistent or they are limited in traditional instruction, such as virtual navigation within inaccessible or dangerous places.

The effectiveness of virtual environments as a complementary tool for learning has been demonstrated in different studies. An empirical study was conducted comparing learning activity in situ of a real environment with a desktop virtual reality environment. The outcomes support the fact that the real environment is superior for the learning activity; however, they also show that the virtual environment is useful for priming and reinforcing curriculum material or for situations where the real environment is inaccessible. Thus, the virtual environment may serve educational goals when used properly, and it can come close to the real environment; but the combination of the experience in a real environment with virtual environments is even more effective in maximizing training, and has greater impact on students (Harrington, 2011). Furthermore, VR systems help to make training faster (Park, Jang & Chai, 2006).
Virtual reality in training within the electrical sector

Regarding the field of electricity, VR has also been applied to training. For instance, VR is among the most innovative approaches used to improve training methods in this field. Among others, we have found applications to teach techniques for maintenance, repair, and diagnosis of complex machinery (Arendarski, Termath & Mecking, 2008); interactive 3D visualization of complex equipment such as transformers and generators; authoring tools, which allow technical specialists in the electricity industry to share and transfer their knowledge and experience to new staff members, where the trainee can interact with the scenario created by the tutor.

As an example of a more specific application in this field, Sun, Chen and Liu (2014) present a virtual reality system to train personnel in manual operation of equipment in an electrical substation. This system enables the operation of a substation by interacting with the switching equipment and performing operations as in the real world. There are many similar VR systems involving 3D electrical substations, most of them devoted to training operators and substation simulators (Barata et al., 2015; Cardoso et al., 2013; Yan, Wang & Zhao, 2012; Wang & Li, 2010).

Animated agents

Extensive research has been conducted on animated agents. The motivation behind an animated agent is, in most cases, to add or strengthen social and communicative features in a pedagogical system (McQuiggan, Mott, & Lester, 2008). These characters increasingly live in digital media, and there is no doubt that they induce affective responses. Works are focused on finding out to what extent these characters influence affective responses to the system, which already exist even without this visualized character. Another interest of research is to know whether affective response patterns known from social psychology can be elicited by manipulating only the static visual appearance of the character, while controlling for voice, content and visual dynamics such as movements and facial expressions (Gulz, Ahlnér & Haake, 2007).

Another strand presents a theoretical framework addressing three aspects of embodied pedagogical agents: static visual appearance, pedagogical role, and communicative style (Haake & Gulz, 2009). This framework was applied in a user study where 90 school children were presented with either an instructor or a learning companion condition. They had to choose between eight visually different embodied pedagogical agents. The outcomes showed that there was a significant tendency to choose a task- and relation- oriented pedagogical agent. The goal of the study was to explore possible relations between the three aspects mentioned above with respect to user preferences.

Affective modelling in virtual agents plays a role in motivating users, supporting them through stressful tasks, and increasing users’ abilities to recognize and regulate emotions (Sabourin & Lester, 2014). Agents can exhibit different strategies; among others they can be persuasive, caring, supportive, and empathic (Woolf et al., 2009). Empathetic behavior has been suggested to be an effective way for animated agents to provide feedback to students and enhance learning. For instance, Moridis and Economides (2012) conducted a survey to examine the impact of agents’ emotional facial expressions and tone of voice combined with empathetic verbal behavior when displayed as feedback to students’ emotions of fear, sadness, and happiness in the context of a self-assessment test. The results of the survey indicate that an agent performing parallel empathy, displaying emotional expressions relevant to the emotional state of the student, may cause this emotion to persist. Moreover, the agent performing parallel and then reactive empathy appeared to be effective in altering an emotional state of fear to a neutral one.

Blended learning

Several definitions of blended learning have been proposed, yet there is no single accepted definition. Pankin, Roberts and Savio (2012) define blended learning as structured opportunities to learn, which use more than one learning or training method, inside or outside the classroom. This definition includes different learning or instructional methods, different delivery methods, different scheduling, and different levels of guidance. Some of the advantages of Blended Learning recurrently reported in literature are: reduction of learning costs, the flexibility of eliminating space and temporal restrictions of learning; this is possible for the asynchronous and synchronous capabilities of BL (Garrison & Vaughan, 2008).
In universities and companies, successful applications of blended learning can be found (Vásquez-Ramírez et al., 2012). Barragán et al. (2009) report positive appraisal in the acquisition of competences by students, as much in the comprehension of the purposes and processes of the learning as in the achievement of the competences planned. They applied an electronic portfolio as a tool to evaluate two courses in the Pedagogy program. Poon (2012) reported a case study regarding the use of blended learning as a delivery method at Nottingham Trent University in the United Kingdom. The author aimed to examine the benefits that blended learning provides to students’ learning experiences. Data collected for this study included interviews with academics and responses from students to a questionnaire survey. The research findings formed the basis of recommendations for the development of learning and teaching practices and approaches that will enhance students’ learning experiences.

Nickel et al. (2015) report a comparative study in which two groups of novice students are trained in Laparoscopic Cholecystectomy (LC) using blended learning (BL), where the VR group completed the LC significantly faster and more often within 80 min than BL. The BL group scored higher than the VR group in the knowledge test and both groups showed equal operative performance of LC. These authors suggest that multi-modality training programs that combine the advantages of both approaches should be developed. On this stance, there is also the term Blended Reality defined as an interactive mixed-reality environment where the physical and the virtual are intimately combined in the service of interaction goals and communication environments (Hoshi et al., 2009). Blended reality is intended to take advantage of the potential in combining virtual worlds and face-to-face classes.

**Roadmap for training**

In order to have efficient training, we have defined a four phase roadmap that takes us from traditional training to a system that allows adaptive, intelligent self-training. This roadmap may be applied in different learning environments, but we want to use the electrical test field to develop the roadmap. The roadmap integrates different technologies, such as: virtual reality, empathetic animated agents, blended learning, and adaptive training. In this section, we present a detailed description of the roadmap and of its main elements.

**Phases of the roadmap**

The difficulty involved in performing electrical tests requires an adaptive training model where the individual state of trainees is taken into consideration and where self-training is also required. We are developing a model and a platform for blended training that intelligently adapts to the needs of the trainee. Figure 1 shows the roadmap we have defined for the training system. Currently, we are working on Phase 3; the VRS has been implanted successfully, now the empathetic agent and trainee model are under development. Our final aim is to have an intelligent system that supports self-training in an adaptive way (Phase 4).

**Phase 1: Traditional training**

Traditional training consists of both classroom lessons and camp training. In the classroom, students learn about theoretical aspects of electricity, safety procedures and tools, materials and equipment management, and theoretical basis for electrical tests. The camp training allows students to practice and complete their knowledge on conducting electrical tests performance. In the beginning, the trainee plays an auxiliary role, helping the expert technician to
conduct such tests. The learner enhances his knowledge progressively through time, until the expert eventually certifies that the student (or auxiliary technician) is fully trained to conduct the tests by himself. This takes an amount of time that may vary from student to student, depending on their personal ability and progress.

Regarding classroom teaching, instructors keep records of student progress in traditional lists of marks. In the case of camp training, students have a type of personalized instructor while helping to perform real electrical tests. Thus, the progress is observed in practice rather than on record. The main drawback here may be found in the time needed to achieve the state in which the student is fully trained, since conducting electrical tests depends on availability of substations.

**Phase 2: Virtual reality supported training**

This stage, “Training supported by Virtual Reality,” means traditional training enhanced by a Virtual Reality System. The system complements the classroom training; it supports the instructor when teaching how to conduct electrical tests. Here, besides in addition to the traditional recording of students’ progress by instructors, the system keeps its own records, which complement the traditional ones.

The system provides students with knowledge before they go to camp training. Using the system, students virtually perform and practice electrical tests. Here students can start learning at their own pace, even if they are not in a programmed course. This may impact the number of students who can learn about electrical test performance.

**Phase 3: Intelligent systems supported training**

In this stage, the VR training systems must be complemented by integrating technologies based on artificial intelligence, in such a way that the records kept by these systems are complemented to include the trainee model. On the other hand, these systems also include an empathetic agent, which also introduces the technology of affective computing within the learning process.

Thus, with all this arsenal of technology, in a system like this, students will be guided step by step, by means of animation, within a 3D virtual substation, throughout the conduction of an entire electrical test. At each step, an animated agent will provide explanations and show emotions empathetically in order to enhance learning. Furthermore, the system will decide what a student must learn according to his progress and based on the trainee model.

**Phase 4: Adaptive and intelligent self-training**

In this last stage training can be supported by the integration of different technologies within the training system. Here, the system includes all the capabilities described in the previous three phases and adds an adaptive behavior based on the student model. Thus, the system may be able to guide and suggest the most suitable topics to be learned by each student according to his progress history. This also allows the system to show an intelligent behavior. In order to achieve this stage, the student model is updated every time a student is using the training system, where progress and failures are recorded, so that the system is able to make correct decisions regarding the selection of topics when guiding students.

**Virtual reality system**

The virtual reality training system is expressly designed to support electrical tests learning on primary equipment of distribution substations (Pérez Ramírez et al., 2014). Conducting electrical tests involves dangerous activities. For instance, most of them require isolating the electrical equipment under testing; this involves opening and closing energized blades. Some tests also require injection of high currents and voltages.
Electrical tests in substations allow foreseeing potential problems that could damage substation equipment and, in turn, originate interruptions, which are the last situation electricity companies expect to face. On the other hand, erroneously performed tests may result in accidents or equipment damage. Thus, efficient training is mandatory in order to ensure optimal operation of the substation.

Overall, electrical tests consist of a sequence of steps where the equipment to be tested must be bypassed so that it is not energized. Then, the testing equipment is connected to the primary equipment under test, and outcomes of the tests are recorded. If no other tests are performed, the testing equipment is disconnected and removed. Finally, the primary equipment under test is connected again and reset. It should be mentioned that safety regulations and measures must be observed in every step (CFE, 2003). Figure 2 shows a transformer under test in the VRS. On the left, the panel shows information and explanations regarding the electrical test in progress; the text is in Spanish as the VRS is used in Mexico.

The VRS shows how to perform an electrical test. A control panel with several controls and instructions provides support to the trainee. The VRS includes 40 electrical tests to different primary equipment such as transformers, interrupters, capacitors bank, and so on. Among the tests, we can mention isolation resistance, power factor, and operation time.

The system has been installed across the country. It is used as a complementary tool to traditional training and it has been used in dozens of training courses. Figure 3 shows trainees using the VRS in a training course for electrical tests.
Thus far, the system has been helpful as a supporting tool to improve training. Nevertheless, other technologies can be integrated to the VRS, so that they can exhibit intelligent and adaptive behavior, which in turn will improve them as training tools. Next section shows the proposal of a model to integrate intelligent and adaptive features to VRS.

**Blended training model**

Currently, we are developing a model and an intelligent system to support traditional training. In this new model the knowledge state of students is recognized in order to have more precise information for instructors to grant a certification or to recommend attendance to other training courses. Moreover, the model includes recognizing the affective state of trainees, and an empathetic agent properly presents the instruction. The training model is presented in Figure 4.

The training and examination materials were developed and designed by a team of experts. Currently, the instructor plans the course. The instructor teaches trainees in the classroom and trainees also learn and practice with the VRS as much as they want. When trainees have attended the appropriate courses, they have to serve as auxiliary electricians in a real substation.

In the current phase (third phase of the roadmap presented in Figure 1), the knowledge of the trainee is not being used to adapt the instruction in the VRS, since the steps in electrical tests have to be performed sequentially; it is used by instructors to adapt the actual course. We propose to model the trainee’s knowledge and affect. This trainee model is useful for instructors, who will be able to design new courses and new examination materials, redesign training material, and certify trainees. The empathetic agent uses the identified affective state of the trainee to show facial expressions in order to motivate him. The affective trainee model is described below.

**Affective trainee model**

Our proposal involves integrating information from the context and from the facial expressions of trainees, in order to obtain a more precise affective state. We include information about causes (context of the training situation and trainees’ characteristics) and observable effects of emotions (facial expressions).
For the causes of emotions, the affective trainee model uses the OCC model (Ortony, Clore & Collins, 1988) to provide a causal assessment of emotions based on contextual information. The OCC model defines emotional state as the outcome of the cognitive appraisal of the current situation with respect to one’s goals. The trainee model consists of a dynamic Bayesian network (DBN) that probabilistically relates personality, goals, and interaction events with affective states, based on the theory defined by the OCC model. Figure 5 shows a high-level representation of the model, where each node in the network is actually a set of nodes in the detailed model. The model is based on the proposal (Conati & Maclaren, 2009) and on our previous work (Hernández, Sucar & Arroyo-Figueroa, 2013).

Figure 5. High-level DBN for the affective student model

The DBN models the dynamic nature of emotions. To infer the affective state at $t_{n+1}$, it considers the student’s knowledge, personality, and the tutorial situation at that time, as well as the student affective state at $t_n$. The tutorial situation is defined based on the results of the trainee’s actions.

The trainee’s appraisal of the current situation given his goal is represented by the relationship between the goals and the tutorial situation nodes through the satisfied goals node. The influence of the appraisal process on the trainee’s affect is represented by the link between the satisfied goals node and the affective state node.

According to the OCC model, one’s goals are fundamental to determine one’s affective state, but asking the students to express these goals during training would be too intrusive. Consequently, the goals in our network are inferred from indirect sources of evidence. We use personality traits as a predictor of the trainee’s goals, but we also include the student’s knowledge.

This representation of the students’ affect is based on the causes of the emotions, but sometimes there is not enough information to identify a precise affective state. Therefore, we are integrating other parameters such as voice and facial expressions. Currently, we have implemented an emotion recognizer through facial expression.

On the side of the effects of emotions, emotions are detected through facial expressions. The method used for the detection of visual emotions is based on Ekman’s theory (Ekman & Friesen, 1978), which recognizes 7 basic emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. The recognition system consists of three components: face detection, facial feature extraction, and emotion classification (Figure 6). To implement face detection and feature extraction we used Java, OpenCV (Open Source Computer Vision), and JavaCV libraries. For emotion classification we used the Neural Network Java Framework Neuroph.
We applied the Haar-like features cascades method (Viola & Jones, 2001), to detect the face, the eyes, and the mouth of the person. At the beginning, when the student’s face is detected, some components are invoked in order to find the mouth, right eye, and left eye. For optimal image processing, a ROI (Region of Interest) method was used, delimiting the search space and discarding the rest of the image. Once the objects in the image are found, a series of transformations are performed, facilitating the search for edges on objects and calculating their opening distances. These data feed input to the neural network for emotion classification. To calculate the opening distance (points) of the mouth, left and right eye, different transformations were performed in regions of interest where the objects are found in the image. These modifications allow the application to perform feature extraction, with optimal performance in image size, in addition to image noise cleaning and handling of certain pixels to identify objects in regions of interest.

![Figure 6. Emotion recognition system](image)

The Gaussian average operator was considered to allow cleaning of the image obtained by the application. Initially, Gaussian \( g \) was used where coordinates \( x, y \) are controlled by the difference \( \sigma^2 \) according to equation 1 (Nixon & Aguado, 2008):

\[
g(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}
\]  

(1)

The results obtained using the Gaussian average operator were smoother images, removal of photo details, allowing a greater focus on large structures. Another influential factor in the images is brightness, which may hinder the transformation process as it does not allow the definition of some edges and structures. So we applied the histogram equalization process where the image obtained passes through a non-linear process, which tends to emphasize the brightness in a particular way to make it more suitable for recognition. With this application, the process makes changes producing an image with a flat histogram, where all levels are similar. Then, for a range of \( M \) levels, the histogram draws the points. For the input (old) and output (new) image, the number of points per level is denoted as \( O(l) \) and \( N(l) \) (for \( O \leq l \leq N \)), respectively (Equation 2).

\[
\sum_{l=0}^{M} O(l) = \sum_{l=0}^{M} N(l)
\]  

(2)

Since the output of the histogram is uniformly smooth, the cumulative histogram to the \( p \) level (for an arbitrarily chosen \( p \) level) should be a fraction of the total sum. Then the number of points in the output image is the ratio of the number of points in the range of levels of the output image (Equation 3).

\[
N(l) = \frac{N^2}{N_{\text{max}} - N_{\text{min}}}
\]  

(3)

The last transformation is a threshold that allows the starting point to be distinguished, to calculate opening distances in order to determine the border points of the mouth, right eye, and left eye. To specify a certain level, pixels are only
set in two colors: white for high level and black for low level. To represent the probability of distribution of the intensity levels, the following equation (equation 4) is used.

\[ P(I) = \frac{N(I)}{N^2} \]  

(4)

We trained the neural network with the RaFD face database (Langner et al., 2010). The RaFD (Radboud Faces Database) is a database of facial expressions that includes a set of 67 models including men, women, boys, and girls. The models show eight emotional expressions corresponding to FACS (Ekman & Friesen, 1978), in three different viewing directions, with five positions of the camera. The size of the database is 8040 different facial expressions. The success rate in recognizing emotions of our algorithm was 85%.

For the time being, we are focused on the development of the animated agent that will use the affective state identified in the trainee to show an empathetic behavior.

**Animated agent**

Training activities are presented to trainees through an animated pedagogical agent. These agents represent a major trend in order to have a more natural human–computer interaction (Breese & Ball, 1998). Animated pedagogical agents interact face-to-face with the students through facial expressions, gaze, emotions, and deictic gestures, and they inhabit learning environments to accompany students. Animated pedagogical agents have a significant impact on training systems as they give the impression that someone is on the other side (Sagae et al., 2012); thus, the trainee perceives a behavior that is different from a traditional system and more like human behavior. Among the behaviors of an animated pedagogical agent are those typical of training systems, but there are some particular behaviors of these characters, such as demonstrations of complex tasks, observing and assisting the trainee in the performance of their tasks, in addition to guiding trainees in virtual spaces (Wang et al., 2008).

To develop the agent, we are using the characteristics of operators, such as wearing uniform and safety helmet, among other features. We believe that by representing the tutor as an electrician, operators will accept the training environment, as it is more natural for electricians and hence useful for learning. Figure 7 shows the animated agent wearing glasses and helmet.

![Figure 7. Animated agent wearing glasses and helmet](image)

As an initial proposal to promote a positive affective state in trainees, we have included empathy in the behavior of the agent. Empathy is the ability to perceive, understand, and experience others’ emotions. This construct has been incorporated in animated agents with the aim to achieve believability, social interaction, and user engagement (Hone, 2006). In this initial phase, the animated agent will deploy the emotions recognized in the student model as described above. In Figure 8, we show the agent integrated into the VRS; he is expressing happiness.
We conducted a study to evaluate the current design of the agent and gather information to refine the agent in his appearance and ability to show emotions. In the next section, we describe the study and main findings.

**Affective animated agent study**

Before having the empathetic agent completely integrated into the VRS, we wanted to know if electricians would welcome the agent. Therefore, we carried out a study presenting the agent showing the basic emotions proposed by Ekman (Ekman & Friesen, 1978). In the study, we wanted to know the expressiveness of the agent. The aim was to find out if electricians would feel identified with the agent, and if he would match their preferences regarding presentation, behavior, and appearance of the agent. In this way, we can design the agent taking into consideration the following three aspects: (i) people’s perception, (ii) psychological models, such as the OCC Model (Ortony, Clore & Collins, 1988), and (iii) the theory proposed by Ekman and Friesen (1978).

In the study, 35 people participated –25% female and 75% male–; on average, they are 36 years old and have 12 years of experience. The participants work at the Mexican electricity utility and they are familiar with electrical testing and with the VRS. Therefore, their feedback is invaluable as they are a representative sample of the people who are intended to use the training system.

We explained participants the aims of the study; then, we introduced the agent showing the eight facial expressions (Figure 9), for the emotions proposed by Ekman (Ekman & Friesen, 1978).

The study consisted in showing the participants the facial expressions of the agent and asking them to identify the emotion represented by each expression. This was carried out applying a five-section study; the first three are the following: (1) Writing down the emotion perceived providing an open answer; that is to say that participants could write any emotion they perceived. (2) Selecting the emotion from a list without going forward or backwards in the study to see other facial expressions or to change answers. And (3) Selecting the emotion from a list; in this case participants could go forward or backwards in the study to see other facial expressions, compare, and change answers if they so desired.
Figure 9. Emotions included in the study

Table 1 shows the answers from the three sections of the study. In section 1, we asked participants to try to identify the emotion in the agent providing an open answer; since the question was open, they mentioned a number of different emotions. This is why the level of agreement between answers of the participants and the emotion intended is lower than in question 2 and question 3. The last column shows the percentage of participants that were consistent in their answers; that is, the participant identified the same emotion in all three questions. In these three questions, we can observe that the emotions of sadness, happiness and surprise were identified in a higher percentage in contrast with the emotions of fear, anger and disgust, which were identified in a lower percentage.

Table 1. Agreement between intended emotion and perceived emotion by the participant

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Open</th>
<th>List changes not allowed</th>
<th>List changes allowed</th>
<th>Consistency in answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sadness</td>
<td>80.0%</td>
<td>100%</td>
<td>94.3%</td>
<td>77.1%</td>
</tr>
<tr>
<td>Happiness</td>
<td>62.9%</td>
<td>94.3%</td>
<td>94.3%</td>
<td>60.0%</td>
</tr>
<tr>
<td>Surprise</td>
<td>60.0%</td>
<td>80.0%</td>
<td>77.1%</td>
<td>42.9%</td>
</tr>
<tr>
<td>Fear</td>
<td>28.6%</td>
<td>40.0%</td>
<td>48.6%</td>
<td>11.4%</td>
</tr>
<tr>
<td>Anger</td>
<td>11.4%</td>
<td>22.9%</td>
<td>20.0%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Disgust</td>
<td>2.9%</td>
<td>37.1%</td>
<td>31.4%</td>
<td>2.9%</td>
</tr>
</tbody>
</table>

(4) Furthermore, we asked participants their opinion on the relationship between learning and affect and its use in training systems. We observed that all the participants consider that the mood or emotional state is important for learning and more than 60% consider that animated agents are useful for learning. Almost 60% would choose a training system with an animated agent. However, regarding the suitability of this particular agent for our VRS, 50% thinks that it is suitable, while the other 50% believe that it is not suitable for the particular VRS and intended trainees.

Finally, (5) we asked participants why they answered the way they did, and we found the following: Most people recognized that emotions play an important role in learning; they also associate emotions with concentration and even they even propose that happiness is a good emotional state, suitable for learning.
Analysis of results and discussion

Although the VRS has shown to be useful in training, ours is not yet adaptable to individual electricians. Namely, the system provides instructional content regardless of the particular states of trainees. Now we want to provide personalized and intelligent support to trainees and instructors. To do so, we defined a four-phase roadmap.

Aiming to improve traditional training, we have developed in phase two a VRS that provides different advantages of the VR technology. One of the main benefits is that the system increases the availability of knowledge about electrical tests to primary equipment in substations. We are currently working on the 3rd phase, whose progress is described in this document. We want to add an empathic agent and the trainee model to the VRS, within a blended learning model. All this, in turn, will be the basis to support the last phase where affective states and knowledge of students will allow enable the development adaptive learning.

In the effort to include empathic agents to this SRV, the study provided useful information to improve the appearance of the agent. The participants agreed that an animated agent would be useful for learning; they claimed that they would try it even for curiosity. Only a few believe that agents can be a distracter, but a good design might avoid this. Regarding the 3D representation of the instructor wearing a mandatory work uniform, most of them agreed that it would make the environment more familiar to new electricians. Opposing opinions focused on deficiencies in the agent, but again these can be overcome by refining its design. Thus, most participants perceive the use of animated agents within a virtual learning environment as positive. They even provided proposals to improve the animated agent, suggested the personalization of the agent depending on specific learning contexts, and also suggested the use of more than one agent, where at least one of them is female. We were surprised that our assumption about the appearance of the agent (wearing uniform) was proven when we did not even explicitly ask for it.

The study also indicated that different emotions, represented as they are now (Figure 9), must be refined. Among the clearest emotions recognized were sadness, happiness, and surprise, but there was some disagreement in recognizing other emotions such as anger and disgust.

On the other hand, neural networks are used to recognize emotions from student’s facial expression. This will support the development of the empathy to be shown by the agents.

Our VRS, as it has been accepted and used as a complementary training tool, and it has been helpful to improve training. We have identified many similar VR developments but these are mostly devoted to simulation and training operators of substations. All these other systems are different from our VRS. Rather than operation, our VRS is devoted to training on electrical tests, which are performed on the substation’s primary equipment. We expect that the addition of empathic agents as well as the detection of the affective and knowledge states, will constitute a step forward to improve our VRS as a learning tool in this third phase of the roadmap.

Conclusions and future work

There may be instructional domains where learners can self-learn using a system whose instructional content is comprehensive and exceptionally well done. In such cases, the presence of an instructor may not be crucial for trainees. Nevertheless, for the system mentioned in this paper, this is not the case. Electrical Test procedures involve high risk and physical activity, elements that are not provided by a non-immersive VR training system. The point here is that these systems are still limited with respect to the actual performance of electrical tests; there are physical actions, such as climbing up a transformer, removing cables, or observing safety regulations within specific circumstances, whose expertise cannot be acquired by using the system, but by doing the actual work. This is why within this domain, which involves high-risk actions, the systems is not entitled to issue a certificate to enable people to perform electrical tests; this must be the responsibility of a human instructor who will have to cover the physical and practical training and verify the skills of the trainees. Thus, the system is a helpful complementary training tool that can be used to enhance traditional training, but it cannot replace it.

With these premises, we have designed a roadmap for training to achieve intelligent, adaptive systems. We have designed a blended learning model that includes a virtual reality system with an empathetic agent. The adaptation is
based on the affective and knowledge state of the trainee. Instruction is presented by the empathetic agent to motivate trainees and improve learning.

In this paper, we present the results of a study conducted to evaluate the design of the empathetic agent and gather knowledge to refine such agent. We have encouraging results, as the electricians welcomed the agent. They found useful to have a virtual companion during training. The results of the study shed some light to refine the facial expressions of the agent and its overall design. In the immediate future work, we consider improving facial expressions and designing other agents—including female agents—, where some of the details, such as helmet and glasses, are changed. In this way, the electrician can select different agents.

In addition to improving the agent, the agenda for future work includes completing the integration of the affective trainee model and evaluating the precision of the identified affective state. Also, we are planning to show the trainee’s knowledge model to trainees as a self-evaluation tool. Self-assessment is one of the meta-cognitive skills, necessary for effective learning. Students need to be able to critically assess their knowledge in order to decide what they need to study. For the time being the trainee’s knowledge model is only being used by instructors.

Acknowledgements

Yasmín Hernandez and Miguel Pérez wish to thank the Electrical Research Institute for supporting and partially funding this work. Ramón Zatarain, Lucía Barrón, and Giner Alor are very grateful to the National Technological Institute of Mexico for supporting this work. Also, this research paper was sponsored by the National Council of Science and Technology (CONACYT), and by the Secretariat of Public Education (SEP) through PRODEP. Also, we all also wish to thank all the participants in the study conducted.

References


Construction of Multi-mode Affective Learning System: Taking Affective Design as an Example

Hao-Chiang Koong Lin¹, Sheng-Hsiung Su¹*, Ching-Ju Chao², Cheng-Yen Hsieh¹ and Shang-Chin Tsai¹

¹Department of Information and Learning Technology, National University of Tainan, Tainan, Taiwan // ²Department of Applied Foreign Languages, Tung Fang Design University, Kaohsiung, Taiwan // koong000@ms39.hinet.net // ice.shsu@gmail.com // chingju@mail.tf.edu.tw // allen01816@hotmail.com // maney100@hotmail.com

*Corresponding author

ABSTRACT

This study aims to design a non-simultaneous distance instruction system with affective computing, which integrates interactive agent technology with the curricular instruction of affective design. The research subjects were 78 students, and prototype assessment and final assessment were adopted to assess the interface and usability of the system. Prototype assessment consisted of heuristic assessment and the system usability scale, while final assessment adopted the triangular cross-validation method: where the questionnaire for user interaction satisfaction, observation, and interviews were used to explore the effect of learning and obtain qualitative and quantitative information for analysis. According to the experimental results, the usability of the non-simultaneous distance instruction system with affective computing was high; the respondents showed high-level satisfaction regarding interaction with the affective learning system; the training game response mechanism of the system could effectively improve the emotion of learning; there was a significant improvement in the effect of learning based on the affective learning system.

Keywords
Affective computing, Affective tutoring system, Text emotion, Facial expression, Skin potential

Introduction

In recent years, the prevalence of smart devices demonstrates a lifestyle associated with computers in modern society. Regarding different types of program, besides functions, humanity is also critical. The most significant difference between human beings, computers, and machines is that human beings have emotion. Therefore, in order to reinforce the humanity of computers and machines, they should approach human emotion in order to provide proper feedback. With the popularity of the internet, non-simultaneous distant instruction gradually replaces traditional instruction. Emotion is part of the key semantic information of interpersonal relationships. Positive emotion leads to more successful learning processes (Ezhilarasi & Minu, 2012; Eyharabide & Amandi, 2012). Instructional systems can automatically recognize a learners’ emotion and provide proper feedback, thus, it is a significantly potential and essential research indicator (Islam, 2013).

With the prevalence of online learning, if computers can recognize learners’ learning emotion and maintain their position emotion, it will result in learning efficiency and effectiveness (Kerr, Rynearson, & Kerr, 2006). In interpersonal interaction, conversation and facial expression are the most direct methods to recognize others’ emotions; however, people can disguise these two emotions. Therefore, aside from introducing semantic and facial expression emotional recognition, this study measures learners’ physical signals as indicators of emotional judgments. Since physical signals are non-volitional physical reactions, the emotional state is the most objective. The introduction of multi-mode Affective Computing can enhance the precision rate of emotional recognition in Affective Tutoring Systems.

This study applies Affective Computing, as developed by the laboratory, to research the findings and development of Intelligent Tutoring Systems. Through developed Affective Tutoring Systems, it expands different dimensions and functions. The purpose is to strengthen learners’ trust and involvement in learning by a more advanced and complete emotion recognition module. By complete and effective assessment, it demonstrates the value and existence of this study.
Literature review

Affective tutoring systems

Changes of technology are rapid and important, and various kinds of digital technologies and techniques are gradually introduced to instructional environments. As the traditional teachers’ one-to-many instruction model changes, learning models have become diverse (Chen, Kao, & Sheu, 2003). ITS (Intelligent Tutoring Systems) means to provide personalized instruction by computer analysis or direct feedback to students. ITS finishes different instructional tasks by simulating teachers. According to different students’ characteristics and states, it indicates various instructional methods. Affective Computing is a research field that deals with issues of emotions and computers. Generally speaking, it is classified into four research levels: emotional recognition, emotional expression, having emotion, and emotional intelligence. Most studies focused on emotional recognition (Picard & Klein, 2002).

Technology introduction reinforces class activities. In addition, teachers can observe the negotiation, communication, cooperation, and interaction among students (Liaw, Chen & Huang, 2008; Infante et al., 2009; Martin, Pastore, & Snider, 2012). However, without an appropriate learning strategy, learning effectiveness will not be as expected (Peng et al., 2009). ATS (Affective Tutoring Systems) means the defection of students’ learning and emotional states to offer proper emotional feedback and regulate students’ learning emotion (Mao & Li, 2010; Hsu, Lin, Lin, & Lin, 2014). ATS is developed upon ITS, and aims to effectively adapt to students’ emotion by simulating human beings (Ammar, Neji, Alimi, & Gouradères, 2010; Lin, Wang, Chao, & Chien, 2012). Although ATS is developed in recent years, it is the first research that adapts to and recognizes emotion. Thus, this study reviewed this study of Picard (Picard, 2000), who proposed a conceptual module that influences learning emotion, and constructed a system that recognizes learners’ emotional state, provides appropriate feedback, and reinforces learners’ learning (Lin, Chen, Sun, & Tsai, 2012; Lin, Hsieh, Loh, & Wang, 2012).

With the Affective Computing technique, computers can recognize human emotions. ATS is considered as a personalized training model. Lin, Tsai, Cheng, Chao, and Su (2014) combined Affective Computing with a webpage system to develop Affective Computing on a webpage to provide adaptive learning for learners. Mao and Li (2010) undertook an investigation into the key factors that influenced students’ satisfaction when using the affective learning system, and found that the factors included the attitude of students, the effect of tutoring, the accuracy of emotion recognition, the quantity of identifiable emotions, instructive action, and the usability of the system. Sarrafzadeh, Alexander, Dadgostar, Fan, and Bigdeli (2008) developed a system targeting math for primary school students. Through the instruction of a lifelike animated agent, the system analyzed the facial expressions of students to identify their emotions, and showed the emotion of the animated agent. By incorporating affective computing into
A smart tutoring system, Ammar et al. (2010) detected and judged facial expressions and computed emotions, in order to enhance the interaction between instruction and learners, develop students’ interest in learning, help them absorb knowledge, and significantly increase the effect of learning through mutual assistance among learners. The framework chart of the affective learning system is as shown in Figure 1.

**Emotional recognition**

*Chinese semantic information*

Emotion is the key semantic information of interpersonal interaction (Ezhilarasi & Minu, 2012). If emotion can be precisely recognized, it will help to make decisions in better way (Lin, Wu, & Hsueh, 2014). The first condition of text emotion recognition is to understand semantic content to acquire precise information. It should be based on natural language processing and semantic analysis (Yan, Bracewell, Ren, & Kuroiwa, 2008). The assumption of document frequency might lower the precise rate of classification, as many terms with high document frequency are usually unused words or unimportant information (Basu & Murthy, 2012). Therefore, when selecting terms by document frequency, other methods are usually adopted. Lu, Lin, Liu, Cruz-Lara, and Hong (2010) proposed the automatic and hierarchical emotional semantic acquisition system, which is highly intensive. Through an independent database, it automatically recognizes the subjects’ semantic emotional responses to the events of sentences. Regarding classification of emotions, there are two principle methods: monitoring and non-monitoring learning (Feldman, 2013). Studies on emotional analysis demonstrated that the importance of the terms, as calculated by TF-IDF weights, is highly effective (Liu, 2012).

**Facial expression**

Facial recognition techniques are commonly applied in daily lives (Gunes & Piccardi, 2007), such as, automatic facial focus, domestic security and recognition, and facial recognition guard systems, which unlock the door of facial recognition. Camera shutters with multi-smiles is another extended technique of facial recognition. Therefore, it demonstrates the common application of facial expressions and potential. According to faces and facial features, there are generally six kinds of facial expressions: joy, anger, sadness, surprise, fear, and disgust (Ekman & Friesen, 1971; Ezhilarasi & Minu, 2012). Metri, Ghropade, and Butalia (2012) constructed a facial emotional recognition system by facial features, as proposed by Ekman, and enhanced the effect of recognition of emotions through physical poses. Figure 2 shows the steps of recognizing facial expressions. Hsu et al. (2014) integrated facial recognition and semantic recognition by multi-mode for Affective Computing to develop an affective tutoring system.

![Figure 2. Steps of facial emotional recognition (Source: Metri, Ghorpade, & Butalia, 2012)](image)

**Skin conductance degree**

Skin conductance is also called Galvanic Skin Response (GSR), and it means the electronic conductance on skin. When people’s emotions change, their body actions, facial expressions, and physical reactions will change accordingly. For instance, with stimulus, the secretion of skin sweat glands will influence changes of the Galvanic Skin Response. We can measure the change of skin potential by Affectiva Q sensor. Through this instrument, we can obtain the figures and analyze the different waves. The peak of the wave means severely positive and negative emotional reactions, as shown in Figure 3.
Research method

Research process and implementation steps

In order to probe into learning effectiveness and system usability of the multi-mode Affective Tutoring System, as introduced in digital art materials, which include network art, dynamic video-audio technology, recording art, software art, and new media art, 78 college students of a national university in southern Taiwan were invited as the research subjects. Their educational levels were college and university to master. By prototype assessment and final assessment, it evaluated the system interface and system usability. Figure 4 shows the system assessment process of this study. The first step is to put forward a system concept model for the design of the prototype system. After the design, the prototype system is assessed. The assessment of the prototype system consists of two parts—the usability scale of the assessment system for average users, and the assessment of experts. The last step is the final assessment, which adopts the triangular cross-validation method. The research analysis is based on questionnaires, observations, and interviews.

Prototype system assessment

- Assessment of average users—System usability scale: To obtain information regarding the subjective feeling of average users during operation of the system, this study adopted the system usability scale (SUS) as an assessment tool. Developed by the British Digital Equipment Co., Ltd. in 1986, the scale was designed to inform enterprises of the general usability of their products and provide a low-cost, reliable, and fast method. It is a
five-level Likert scale, with each item including five options ranked in an ascending manner: (1) Strongly disagree; (2) Disagree; (3) Average; (4) Agree; (5) Strongly agree. The scale comprises 10 items and adopts the forward and backward cross-questioning method. Usually, respondents complete the scale without discussion after operating the system. Through a formula, scoring will be converted into one with “100 points” as the full mark, where a higher score indicates stronger satisfaction of respondents for the system.

- Assessment of experts--Heuristic assessment: Heuristic assessment is an informal usability testing method used to detect problems regarding usability in the design of a user interface, in order that these problems can be regarded as a focused part of the redesign. In the heuristic assessment, experts follow a group of given usability heuristics, and evaluate the constituents of a respondent interface to see if these constituents are consistent with these heuristics (Nielsen, 1994). Meanwhile, Nielsen suggested inviting 3-5 evaluators, and predicted about 75% of the problems regarding usability.

- Final assessment—Triangulation: As a method of testing research information, triangulation adopts more than two resources to obtain full understanding, and demonstrate a specific reference point or topic, with the aim of enhancing the rigorousness and reliability of research. In general, it is recommendable to include three information sources, which allows an evaluation with diverse perspectives, and provides a neutral stance when two views are contradictory.

Affective tutoring systems

The system framework of this study is as shown in Figure 5:

![System framework](image)

**Figure 5. System framework**

*Emotional recognition module*

- **Semantic recognition module**: By dialogue between the subjects and the interactive agent, this study conducts Chinese semantic emotional recognition; and through the dialogue input by the subjects, the subjects’ emotional state is immediately recognized. The construction is as shown below: (1) construction of emotional dictionary; (2) semantic structure message processing: word segmentation rules, semantic structure, message processing, keywords matching, and word segmentation; (3) acquisition of semantic emotion. Figure 6 shows the process of semantic analysis.
Facial expression recognition module: By open library--EmguCV, this study develops a facial expression recognition module. EmguCV is an OpenCV component packaged by C# for the development of a Visual Studio. EmguCV not only has powerful image processing capacity, but is also an open and free library, thus, system development is less difficult. The steps of facial expression recognition are shown as follows: (1) recognition of human beings’ facial positions; (2) recognized facial features are compared to six kinds of facial expressions by classification - HaarTraining and classification; (3) recognized facial expressions refer to emotions; in the opposition situation, there is no emotion.

Physiological signal module: By Q sensor, this study includes the skin conductance information of the physical signals in the system. Through a Bluetooth connection, Q sensor sets the device as a COM Port. By Visual Studio C#, it obtains related figures. Figure 7 shows the process of physiological signal.

Brainwave concentration and relaxation training module: By NeuroSky MindWave Mobile, this study conducts brainwave concentration and relaxation training. When the subjects learn by this system, they rely on the emotional recognition of the previous three kinds of emotional recognition modules. When the subjects have negative emotion, the system will automatically accumulate the emotion. When a negative emotion unexpectedly occurs, the system will automatically stop the teaching material and record the learning time in the
database. It then activates the Brainwave Visualizer, as developed by NeuroSky, for the subjects’ concentration and relaxation training. With training effectiveness, the subjects can continue learning.

- **System interface**: In this study, the system interface design is divided into 7 zones: function, teaching material, interactive agent, semantic dialogue, facial expression, skin conductance signal, and system record. Figure 8 shows the layout of the system interface.

![Figure 8. Layout of system interface](image)

(a) Tools (see Figure 8 a): video control, which can be played, paused, and stopped; after-class questionnaires include learning effectiveness, scale of system usability, and user interaction satisfaction questionnaires. After learning the subjects, there can be after-class assessment and system assessment; regarding parameter settings, the subjects can regulate facial expression recognition parameters. According to individual conditions, it regulates recognition sensitivity. There are six kinds of recognition, and each kind includes two regulation parameters; the description column provides system instructions, system history, and laboratory introduction. The end button of this system is termination of the course. Three questionnaires will be completed to finish the experiment.

(b) Teaching material (see Figure 8 b): the video playing of teaching material. The videos, as recorded by a teacher in advance, are constructed for the subjects’ learning. Video of this experiment is based on system design. Therefore, the affective design is treated as the instructional content. A combination of instruction content and system allows the subjects to have profound learning experiences.

(c) System record (see Figure 8 c): records all activities of this system, including system starting time, change of course playing, database connection records, results of facial expression recognition, semantic and emotional recognition, Q sensor connection, and accumulation of negative emotion. The recording format is “hour: minute: second. Milli-second recording”.

(d) Facial expression (see Figure 8 d): the subjects’ facial expressions are captured by webcam. When the face and facial expression are captured, a square frame will be drawn. In the figure, the white frame is the subjects’ captured face; the yellow frame is the subjects’ facial expression recognized as joy.

(e) Skin conductance (see Figure 8 e): the subjects’ skin conductance is captured by Q sensor. Captured frequency is 32 times/every second. The system shows the total in the system and conducts emotional sensing.

(f) Interactive agent (see Figure 8 f): the agent designed by this study: batman. According to the emotion recognized by this system, it transfers the movement of facial expressions and interacts with the subjects in a semantic dialogue zone.
Semantic dialogue (see Figure 8 g): the subjects can interact with the interactive agent in this zone. The system will conduct semantic recognition according to sentences input by subjects, and provide immediate feedback.

Researcher’s operation (see Figure 8 h): records the subjects’ time of system learning and current time. Two buttons are function module controls. The researcher assists with the operation button.

- **Learning process:** First of all, interactive agent module: this module is the bridge between a learning system and the subjects. By an agent mechanism, the system can interact with the subjects and properly provide feedback. Second, video course teaching material module: this module is the video course teaching material recorded by a teacher in advance. It is based on video, and the subjects learn online, as in a classroom. Third, learning state recording module: this module is the core of this system. The system automatically and completely saves the subjects’ learning stages and uploads it to a database. Thus, the teacher can immediately recognize each subject’s learning condition, and actively solve problems for the subjects after class in order to reinforce learning effectiveness.

**Experimental results and analysis**

To determine if this non-simultaneous distance instruction system with affective computing could improve the effect of learning, and evaluate the usability of the system, this study invited 78 undergraduates and graduates to participate in the experiment. The respondents were divided into two groups—the Experimental group, which used the affective learning system, and the Control group, which adopted the online webpage learning system. The course materials were videos pre-recorded by the teachers of a university in Kaohsiung, Taiwan. Figure 9 shows the flow chart of the experiment. Prior to learning, the respondents received a learning effectiveness assessment (pre-learning test), and then watched a 16 minute instructive video. After learning, the respondents received another learning effectiveness assessment (post-learning test), completed the questionnaire for user interaction satisfaction, and took the interview. The entire experiment was recorded, and final assessment--triangulation was conducted after the experiment.

![Figure 9. Flow chart of the experiment](image-url)
The system counted the negative emotions captured in the learning of the Experimental group, and the information was added into the learning status database. When the negative emotion reaches a certain level, the system will suspend the course and start a brainwave concentration and relaxation training game. In concentration training, stronger concentration will result in the explosion of a bucket; while sustainable and strong concentration will maintain the explosion of the bucket, and the best explosion time will appear on the interface, as shown in Figure 10. In relaxation training, if the respondents feel more relaxed, a balloon will start to rise; if the respondents maintain a high-level of relaxation, the balloon will float above and rise higher, and the maximum height the balloon reaches will appear on the interface, as shown in Figure 11. The respondents can decide if they want to continue brainwave concentration and relaxation training; if they refuse to continue the training, they can shift the interface to the affective learning system to continue learning.
After learning, both the Experimental group and the Control group immediately received the learning effectiveness assessment (post-learning test), and then completed the questionnaire for user interaction satisfaction. After completing the questionnaire, the Control group finished its task in the experiment, while the Experimental group finished its task only after the interview.

System usability analysis

At the prototype system development phase, this study invites 30 users for prototype system usability analysis. Cronbach’s α acquired by the scale of system usability is .791, and the least reliability of .7 is accepted by this study, as it demonstrates that the reliability of the questionnaires is acceptable. After transforming reverse responses into positive responses in the scale of system usability, this study conducts item analysis, as shown in Table 1. Noticeably, regarding Q4 and Q10, only 46.6% and 66.8% users, respectively, suggest that they do not need assistance, and the subjects indicate that they can use the system without prior knowledge. Hence, the system interface must be simplified for the subjects’ ease of use.

The researcher analyzed the questionnaire filled by the respondents and converted the backward questions in the system usability scale into forward ones for analysis. According to the 5-point scale, the researcher selected the two highest score -- the percentages of “4” and “5” for aggregation analysis. As is shown in the following table, 56.7% of the respondents to Q1 were willing to use the affective learning system on a regular basis; 70.2% of the respondents to Q2 did not believe that the system was too complicated; 83.3% of the respondents to Q3 thought that the system was easy to use; 46.6% of the respondents to Q4 believed that the system entailed little assistance from technicians; 73.4% of the respondents to Q5 thought that all the functions of the system were well integrated; 83.3% of the respondents to Q6 did not believe that the system was inconsistent; 86.6% of the respondents to Q7 thought that most people would master the skills of using the system within a short time; 96.7% of the respondents to Q8 did not believe that the system was too difficult to use; 96.7% of the respondents to Q9 were confident that they could use the system; 66.8% of the respondents to Q10 did not think it necessary to acquire much knowledge to use the system.

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>3.73</td>
<td>.740</td>
<td>0</td>
<td>0</td>
<td>43.3</td>
<td>40.0</td>
<td>16.7</td>
</tr>
<tr>
<td>Q2</td>
<td>3.80</td>
<td>.925</td>
<td>3.3</td>
<td>3.3</td>
<td>23.2</td>
<td>50.0</td>
<td>20.2</td>
</tr>
<tr>
<td>Q3</td>
<td>4.13</td>
<td>.681</td>
<td>0</td>
<td>0</td>
<td>16.7</td>
<td>53.3</td>
<td>30.0</td>
</tr>
<tr>
<td>Q4</td>
<td>3.20</td>
<td>1.064</td>
<td>6.6</td>
<td>20.2</td>
<td>26.6</td>
<td>40.0</td>
<td>6.6</td>
</tr>
<tr>
<td>Q5</td>
<td>3.87</td>
<td>.629</td>
<td>0</td>
<td>0</td>
<td>26.6</td>
<td>60.0</td>
<td>13.4</td>
</tr>
<tr>
<td>Q6</td>
<td>4.10</td>
<td>.759</td>
<td>0</td>
<td>3.3</td>
<td>13.4</td>
<td>53.3</td>
<td>30.0</td>
</tr>
<tr>
<td>Q7</td>
<td>4.20</td>
<td>.664</td>
<td>0</td>
<td>0</td>
<td>13.4</td>
<td>53.3</td>
<td>33.3</td>
</tr>
<tr>
<td>Q8</td>
<td>4.40</td>
<td>.563</td>
<td>0</td>
<td>0</td>
<td>3.3</td>
<td>53.3</td>
<td>43.4</td>
</tr>
<tr>
<td>Q9</td>
<td>4.27</td>
<td>.521</td>
<td>0</td>
<td>0</td>
<td>3.3</td>
<td>66.7</td>
<td>30.0</td>
</tr>
<tr>
<td>Q10</td>
<td>3.80</td>
<td>.847</td>
<td>0</td>
<td>6.6</td>
<td>26.6</td>
<td>46.6</td>
<td>20.2</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>3.95</td>
<td>.7393</td>
<td>.99</td>
<td>3.34</td>
<td>19.64</td>
<td>51.65</td>
<td>24.38</td>
</tr>
</tbody>
</table>

The scores in the scale were obtained according to the scoring of the system usability scale. They reflected the comprehensive assessment of the respondents on the system usability and can be used for the comparison of usability among different versions of the system. The total score (ranging from 0 to 100 points) of the questionnaire was obtained according to the scoring of the system usability: (1) the score of the items labeled with odd numbers was obtained by subtracted 1 from the original score; (2) the score of the items labeled with even numbers was obtained by subtracted 5 from the original score; (3) the total score (ranging from 0 to 100 points) of the questionnaire was obtained by first aggregating the scores of all items and then multiplying the aggregated score with 2.5. Based on the scoring of the system usability scale, the researcher conducted statistics of questionnaire scores of the subjects. The result is as shown in Table 2; where the mean is 73.75, the median is 70.00, the mode is 67.50, the standard deviation is 11.14, and the minimum and maximum are 47.50 and 95 respectively. The score for the system usability is 70, which indicates that most of the respondents were satisfied with the system usability. The average score for the system is “73.75 points,” which shows that average respondents were satisfied with the usability of system. Meanwhile, it was compared with Figure 12, which fell into the zones of “good” and “excellent,” respectively.
Table 2. Conversion results of scores in the system usability scale

<table>
<thead>
<tr>
<th>Number of samples</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>73.75</td>
<td>70.00</td>
<td>67.50</td>
<td>11.14</td>
<td>47.50</td>
<td>95.00</td>
</tr>
</tbody>
</table>

Figure 12. Score distribution of the system usability (Source: Bangor, Kortum, & Miller, 2009)

User interaction satisfaction analysis

The questionnaire for user interaction satisfaction includes 6 dimensions: total use reaction, display of screen, terms and system information, learning, system function and usability, and user interface. Each dimension includes 2~6 items. According to the subjects’ satisfaction, the rating is from 1~7. There are a total of 27 items. Cronbach’s α of the user interaction satisfaction questionnaire is .949, and the least reliability of .9 accepted by this study. Thus, the reliability of questionnaires results is good.

This study analyzes questionnaires responded by the subjects, and obtains the mean by the total scores of the items of the dimensions divided by the number of items. It then conducts descriptive statistical analysis on the results, with the analytical outcome as shown in Table 3. The means of the dimensions of the Experimental group are higher than those of the Control group. Satisfaction with dimensions is at least 5 in the 7-point scale, meaning that the subjects’ subjective satisfaction with the design of the human-machine interface of the system is good.

Regarding mean, we compare satisfaction with the dimensions of the Experimental and Control groups. Satisfaction with using Affective Tutoring Systems is higher than online webpage learning systems, which shows that, in the same video material learning, satisfaction with Affective Computing recognition is higher than non-simultaneous distant instructional systems. In addition, regarding the means of total use reaction, usability, and user interface, the Experimental group is more significant than the Control group, which shows that users’ interaction satisfaction with the system is high.

Table 3. User interaction satisfaction questionnaire--descriptive statistics

<table>
<thead>
<tr>
<th>Dimensions of questionnaires</th>
<th>Experimental group (30 people)</th>
<th>Control group (48 people)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Standard deviation</td>
<td>Mean Standard deviation</td>
</tr>
<tr>
<td>Total usage reaction</td>
<td>5.14 .83</td>
<td>4.40 .75</td>
</tr>
<tr>
<td>Display of screen</td>
<td>5.76 .92</td>
<td>5.09 .89</td>
</tr>
<tr>
<td>Term and system information</td>
<td>5.29 .69</td>
<td>4.78 .99</td>
</tr>
<tr>
<td>Learning</td>
<td>5.49 .83</td>
<td>5.00 .88</td>
</tr>
<tr>
<td>System function</td>
<td>5.37 .99</td>
<td>5.02 1.08</td>
</tr>
<tr>
<td>Usability and user interface</td>
<td>5.38 .89</td>
<td>4.81 .89</td>
</tr>
<tr>
<td>Average</td>
<td>5.405 .73</td>
<td>4.852 .74</td>
</tr>
</tbody>
</table>

According to questionnaire results and t testing of independent samples, this study attempts to determine if satisfaction is different between the Experimental group and the Control group. Analytical results are as shown in Table 4. Assessment results of user interaction satisfaction is Experimental group > Control group and significance .002 < .05, meaning that the satisfaction of the two groups’ is significantly different. Based on previous results, the Experimental and Control groups have significantly different satisfaction using the Affective Tutoring
System and online webpage learning system. Interaction satisfaction with Affective Tutoring Systems is higher than online webpage learning systems.

| Table 4. User interaction satisfaction questionnaires—t-test of independent samples |
|---------------------------------|----------|----------|----------------|-----------------|-----------------|
| Group                           | Number of samples | Mean     | Standard deviation | t value         | Significance (two-tailed) |          |
| Experimental group              | 30       | 5.405    | .73              | -3.238          | .002**           |          |
| Control group                   | 48       | 4.852    | .74              | 3.047           | .003             |          |

Note. **p < .01.

Analysis of learning effectiveness

Learning effectiveness before the experiment

To determine the difference in knowledge between the Experimental group and the Control group, the researcher undertook independent sample t testing on the pre-learning scores of the respondents. As shown in Table 5, significance (.38 > .05) indicates that there was no significant difference between the two groups before learning, meaning that both groups shared similar knowledge before the experiment.

| Table 5. One-way ANOVA of learning effectiveness before the experiment |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Group                           | Number of samples | Mean     | Standard deviation | t value         | Significance (two-tailed) |          |
| Experimental group              | 30       | 47.00    | 16.43           | .878            | .383             |          |
| Control group                   | 48       | 50.21    | 15.23           | 5.047           | .003             |          |

Learning effectiveness after the experiment

The researcher undertook one-way ANOVA of the post-learning scores of the respondents to determine if there was any significant difference between the two groups, and the result is as shown in Tables 6 and 7. The average post-learning scores of the Experimental group and the Control group were 73.00 and 63.13, respectively, which is higher than the pre-learning average score of 47.00 and 50.21, respectively. The significance (.019 < .05) of the post-learning score manifests that the instruction played a significant role in the improvement of learning effectiveness.

| Table 6. Descriptive statistics of the post-experiment learning effectiveness |
|---------------------------------|-----------------|----------|-----------------|          |
| Group                           | Number of samples | Mean     | Standard deviation |          |
| Experimental group              | 30       | 73.00    | 13.17           |          |
| Control group                   | 48       | 63.13    | 20.02           |          |
| Total                           | 78       | 66.92    | 18.26           |          |

| Table 7. One-way ANOVA of the post-experiment learning effectiveness |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Source of variance              | Quadratic sum   | Freedom     | Mean quadratic sum   | F      | Significance |          |
| Inter-group                     | 1800.288       | 1            | 1800.288         | 5.734  | .019*        |          |
| In the group                    | 23861.250      | 76           | 313.964          |        |              |          |
| Total                           | 25661.538      | 77           | 18.26            |        |              |          |

Note. *p < .05.

The researcher subtracted the pre-learning score from the post-learning score, and conducted percentage conversion for independent sample t testing, and the result is as shown in Table 8. The mean of the increase in the learning effectiveness of the Experimental group is 94.94%, whereas, that of the Control group is 29.60%. The significance of (.033 < .05) means that there was significant increase in the learning effectiveness of the two groups, which indicates that the increase in the learning effectiveness of the Experimental group was greater than that of the Control group, and the respondents could improve their learning effectiveness with the affective learning system.
### Table 8. Comparison of the increase in learning effectiveness

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of samples</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Standard error</th>
<th>t value</th>
<th>Significance (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>30</td>
<td>94.94</td>
<td>158.09</td>
<td>28.86</td>
<td>-2.228</td>
<td>.033*</td>
</tr>
<tr>
<td>Control group</td>
<td>48</td>
<td>29.60</td>
<td>35.98</td>
<td>5.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p* < .05.

### Results of the interview

After the experiment, the respondents took an interview that lasted for 3 to 5 minutes. The researcher wrote the respondents’ answers to the above interview questions, which have the following results:

- Most respondents would first turn to their partners for help, and then their teachers, when encountering problems in learning. The majority of the respondents thought it was difficult to communicate with teachers, and thus, would not regard teachers as the first choice when they sought help. Only a few respondents did not turn to teachers or partners for help; instead, they sought help on the Internet.

- Most respondents thought that active help from teachers would facilitate their learning; however, only a limited number of teachers can notice the problems of students and voluntarily offer help in the current educational environment. A few respondents believed that they would not seek help until they face problems, and that active help from teachers might upset or embarrass them.

- Most respondents thought it necessary to keep the mechanism of Question (2). According to them, not all students would actively seek help, thus, the mechanism was good. A handful of students still preferred to seek help on the Internet, as they thought such a method could improve their learning and memory.

- Most respondents thought that the response from the interactive agent was too limited, and that an interactive agent would give the same response to similar emotion.

- Most respondents believed that the brainwave training game was interesting; however, some thought the game was so difficult that they could not fulfill the objectives, and thus, felt frustrated; while others were deeply fascinated by the training, and hoped to break the record kept by themselves or others.

- Most respondents thought that what was taught was boring, and such a mechanism could help them refocus on what was to be taught later. Only a few respondents believed that the game interrupted their learning, was of little help in their learning, and they felt reluctant to continue learning once they started to play the game.

According to the results of the interviews, most respondents were satisfied with the affective learning system, and thought that the learning mechanism, as proposed by the researcher, was effective and necessary. In emotion recognition, the number of keywords for emotion recognition of Chinese meaning was too small, and the system failed to give responses to many daily expressions; the recognition of facial expression was too sensitive; the information about facial expressions used for classification was inadequate, and errors were likely to occur when the system judged a face for the first time. One respondent thought that the interactive agent distracted his attention during learning, and that the brainwave game would interrupt learning.

### Conclusion and future studies

This study incorporated affective computing into a non-simultaneous distance instruction system, and adopted the multi-mode affective perception module to detect the emotions of the respondents during learning. The curricular collocation system design used affective design as the learning material to provide respondents with more impressive learning experiences. Prototype assessment and final assessment were employed to discuss the usability of the system, as well as satisfaction with the system. The former involved heuristic assessment based on expert evaluation, where a system usability scale was used to analyze the usability of the revised system; the latter adopted triangulation, where methods such as questionnaires for user interaction satisfaction, and observations, as well as interviews, were used to explore satisfaction with the system; the researcher designed the assessment of learning effectiveness according to the learning material in order to analyze the learning effectiveness of the respondents. Based on the experiment results, the researcher have come to the following conclusions according to the objective and topic of this study:
• With the non-simultaneous distance instruction system, the researcher integrated affective computing with the instruction of a video course, and developed the affective learning system. The usability of the system scored 73.75, meaning that average users were satisfied with the usability of the system. Therefore, the researcher believes that the usability of the non-simultaneous distance instruction system with affective computing was high.

• This study used triangulation to investigate respondents’ satisfaction with the affective learning system. In the interviews, most respondents said that the system was simple and easy to operate, the interactive agent mechanism made learning interesting, and enhanced their passion for learning. Hence, the researcher thinks that most of the respondents were satisfied with the affective learning system.

• According to the results of affective computing, the researcher collected the negative emotions of the respondents, and started the training game at an appropriate time in order that the respondents could change their emotion during learning. Most of the respondents believed that such a mechanism was helpful to learning at a later stage, thus, the researcher believes that the training game in this system can effectively improve emotion during learning.

• This study aims to develop an affective learning system based on a non-simultaneous distance instruction system. To determine if the system could maintain and enhance the learning effectiveness of the existing system, the researcher divided the respondents into two groups, namely, the Experimental group and the Control group, and conducted simultaneous assessment of learning effectiveness. According to the experiment, the affective learning system not only maintained the existing learning effectiveness, but also improved the learning effectiveness of most respondents. Therefore, this research believes that the affective learning system can effectively enhance learning effectiveness.

The research results and conclusions show the possibility of the popularization of this affective learning system, and offers suggestions for relevant future research.

Suggestions for the affective learning system

During the experiments, the researcher found that the response mechanism of the interactive agent was a key factor that could effectively increase the respondents’ use of the system. If the recognition and accuracy of the existing affective module are enhanced, respondents will feel a stronger intention to use the system, and will not feel bored in the later stages of learning. Regarding the Chinese meaning emotion recognition module, the researcher thinks that more daily expressions can be added to strengthen the understanding of the system, and that a diverse response mechanism can be added to lengthen the interaction between respondents and the interactive agent. Meanwhile, a large quantity of daily life facial expressions should be collected in order to enrich the database, which in turn will enhance the accuracy of real-time facial expression recognition. Regarding the physiological signal recognition module, the researchers think that, if it is impossible to buy specialized devices, such as an affective lab, the physiological signal recognition module should be removed in order to promote the efficacy of the entire system. Regarding the brainwave concentration and relaxation module, the researcher believes that both concentration and relaxation are important factors that influence learning emotion, and that if the two are incorporated into the real-time detection of the system, it will significantly facilitate learning. With non-simultaneous distance instruction concepts, this study adopted video-audio materials for instruction. It is suggested that future researchers make full use of the functions of the non-simultaneous distance instruction system, and add curricular interaction modules, such as a discussion and curricular interaction sections in order to enhance the interaction between teachers and students, stimulate an actual instruction environment, and strengthen the on-site experience.

Suggestions for video-audio course material

The video-audio course material in this study was the PPT of the textbook instructed by teachers, but without the reality of face-to-face instruction; hence, the researcher thinks that interactive materials, such as augmented reality, should be added to involve respondents in the instruction and reduce boredom. The researcher believes this affective learning system will attract more users. Additionally, pictures of teacher instruction can be added into the video-audio course section to enhance the on-site experience. Regarding the assessment of learning effectiveness, the researcher suggests that future researchers follow the mode, as this study involved the demonstrations of experts, which indicate a high-level of reliability and validity.
Suggestions for the interactive agent

The interactive agent of this study is a singular role, thus, the researcher suggests that a paper doll system be added by future research in order that respondents will have an alternative. Meanwhile, additional accessories for the interactive agent can be added to enhance respondents’ identification with the interactive agent, and thus, improve their learning experience. In terms of the emotion of the interactive agent, more emotional behaviors, such as jumping with pleasure, can be added to reinforce the visual experience.

References


A Review on the Use of Robots in Education and Young Children

Lai Poh Emily Toh1, Albert Causo1*, Pei-Wen Tzuo2, I-Ming Chen1 and Song Huat Yeo1

1Robotics Research Centre, School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore // 2National Institute of Education, Singapore // emilytohlp@ntu.edu.sg // acauso@ntu.edu.sg // peiwen.tzuo@nie.edu.sg // michen@ntu.edu.sg // myeosh@ntu.edu.sg

*Corresponding author

(Submitted August 13, 2014; Revised February 24, 2015; Accepted August 27, 2015)

ABSTRACT
A systematic review was carried out to examine the use of robots in early childhood and lower level education. The paper synthesizes the findings of research studies carried out in the last ten years and looks at the influence of robots on children and education. Four major factors are examined – the type of studies conducted, the influence of robots on children’s behaviour and development, the perception of stakeholders (parents, children and educators) on educational robots, and finally, the reaction of children on robot design or appearance. This review presents the approach taken by researchers in validating their use of robots including non-experimental (mixed-method, anecdotal, cross-sectional, longitudinal, correlational, and case studies) and quasi-experimental (pre- and post-test). The paper also shows that robot’s influence on children’s skills development could be grouped into four major categories: cognitive, conceptual, language and social (collaborative) skills. Mixed results are shown when it comes to parents’ perception of the use of robots in their children’s education while design was shown to influence children’s perception of the robot’s character or capabilities. A total of 27 out of 369 articles were reviewed based on several criteria.

Keywords
Early childhood education, Lower education, Educational robots, Review

Introduction

With the rapid development of technology in the 21st century, the use of multi-media tool in education has become increasingly popular. Notwithstanding their usual engineering applications, robots are being used more in schools. According to Beran et al. (2011), children are also playing more with technologically advanced devices during their playtime. Subsequently, studies were conducted to investigate robot use’s influence on children’s cognition, language, interaction, social and moral development (Wei et al., 2011; Kozima & Nakagawa, 2007; Shimada, Kanda & Koizumi, 2012; Kahn et al., 2012). Recent studies (Wei, Hung, Lee & Chen, 2011; Highfield, 2010; Chen, Quadir & Teng, 2011) reported that robot use encourages interactive learning, making children more engaged in their learning activities. This increase research on robot application to education needs systematic look at the direction taken this past decade in order to elucidate a roadmap for future studies.

Recent reviews on the use of robots in education show the challenges faced by researchers in this field. Benitti (2012) points out that more than 70 papers could have qualified in his review work but only 10 provided quantitative measurement on the use of robots in education. From these ten papers, only those that discuss the potential of using robots in all level of education and highlight the non-engineering benefits were selected.

Mubin et al. (2013) analysed research works from through the actual robots used. The major factors identified were robot’s role, type (physical form), behaviour (capabilities and interaction capacity), learning activity type, and venue (inside or outside of classroom) where learning takes place. Mubin et al. (2013) and Benitti (2012) find similarity on the topics where robots were being used in education – learning language, science, and technology. Although Mubin et al. (2013) differs by pointing out the various roles played by the robot in education – as tutor, tool, or peer.

The reviews provide good starting points for researchers, the criteria (Benitti, 2012) and perspective (Mubin et al., 2013) taken by these two papers could potentially miss those that could be relevant to researchers in the field. Moreover, other factors critical in the use of robot in education may have been overlooked, like the effect of design on interaction or the importance of parent’s perception in the success of implementing a robot-in-education project.

The aim of this paper is to assess the effectiveness of using robots in studies published within the last decade. We look at effectiveness as having four sub-factors – the study type done by the researcher, the influence of the robots on
the behaviour and development of students, the perception of stakeholders (parents, educators and children) about the robots, and the importance of design or robot appearance. To achieve this aim, we would focus on articles on the application of robots in early childhood and lower level education and evidence for the factors would be analysed.

The rest of the paper is organized as follows. The review approach, especially the search and selection strategies, is discussed in details in the next section. The discussions on the four factors above are described in the succeeding sections. The conclusion provides a summary and presents the remaining challenges in this research field.

**Review approach**

To limit the papers to be reviewed, we implemented a search and selection strategy using specific keywords in electronic databases. We started with 369 articles and narrowed it down to 27.

**Search strategy**

Articles reviewed were limited to those published in English from 2003-2013. To gather as many papers as possible, five major databases were searched: IEEE Xplore, Academic Search Premier, ERIC (Educational Resources Information Center, ScienceDirect, and SpringerLink. Only articles published in journals have been included for review, with some exceptions.

Initially, search terms like “robots” and “education” was keyed in but in order to narrow down the result, we used a similar approach to what Benitti (2012) employed. Table 1 shows the five databases and the keywords used for each one.

<table>
<thead>
<tr>
<th>Database</th>
<th>Search protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE Explore</td>
<td>(((&quot;robots&quot;) AND &quot;education&quot;) AND &quot;learning&quot;) AND &quot; teaching&quot;) AND &quot;robotic&quot;) under advanced search options &lt; Journal &amp; Magazines&gt;, &lt; Publication Year : 2003-2013&gt;, &lt; Full Text and Metadata&gt;</td>
</tr>
<tr>
<td>Academic Search</td>
<td>“Robots” AND “Education” AND “Learning” Search &lt;Full Text&gt;, &lt;Date Published: 2003 to 2013&gt;, &lt;Peer Reviewed Scholarly Journal&gt;</td>
</tr>
<tr>
<td>ERIC</td>
<td>“Robots” AND “Education” AND “Learning” Search &lt; Full Text&gt; &lt;Peer reviewed&gt; &lt;Journal&gt;, &lt;Date Published: from 2003 to 2013&gt;</td>
</tr>
<tr>
<td>Science Direct</td>
<td>Search Terms: ‘Robots’ AND ‘Education’ AND ‘child’ and ‘learning’ AND LIMIT-To (topics, “child, robot”) AND LIMIT-To (Topics, “child, robot”), &lt;Date Published: Year: 2003 to 2013&gt;</td>
</tr>
<tr>
<td>Springer Link</td>
<td>Search Terms: “education” AND “robots”, Search under: &lt;Education and Language&gt;, &lt;Learning and Instruction&gt;</td>
</tr>
</tbody>
</table>

**Selection strategy**

This review focuses on articles that reported the use of robots in early childhood education. Selected studies were relevant from early to secondary education context and focused on robot or robotics influence on learning, pedagogical and developmental domains. The studies selected should report the use of robots as an educational tool.

Given the broad inclusion criteria, we managed to find 369 articles in all (see Table 2). To further narrow down the scope of the review, the following exclusion criteria have been implemented:

- **Exclusion Criteria E1**: Article reported the technical use of robots, designs or innovations.
- **Exclusion Criteria E2**: Article reported robotics as a teaching subject.
- **Exclusion Criteria E3**: Article reported studies conducted in higher or university education.
- **Exclusion Criteria E4**: Article reported the use of robots as assistive technologies.
- **Exclusion Criteria E5**: Article did not mention on the use of robots in education.
As shown in Table 2, with the above exclusion parameters, only 27 papers were left. A large number of papers were excluded due to the focus on robots or robotics as the teaching subject (a total of 132 articles based on E2). Most of the engineering articles excluded mentioned the use of robot in education in passing or as a justification for its design; 115 articles were removed based on E1. Moreover, around 12% of articles were excluded because robots were reported as an educational tool for higher education.

<table>
<thead>
<tr>
<th>Database</th>
<th>Selected articles</th>
<th>Total reviewed</th>
<th>Excluded criteria articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE Explore</td>
<td>11*</td>
<td>59</td>
<td>E1: 16 E2: 15 E3: 10 E4: 3 E5: 4</td>
</tr>
<tr>
<td>Academic Search Premier</td>
<td>5</td>
<td>188</td>
<td>E1: 79 E2: 70 E3: 25 E4: 9 E5: 0</td>
</tr>
<tr>
<td>ERIC</td>
<td>4*</td>
<td>10</td>
<td>E1: 0 E2: 3 E3: 0 E4: 0 E5: 3</td>
</tr>
<tr>
<td>Science Direct</td>
<td>3</td>
<td>46</td>
<td>E1: 14 E2: 0 E3: 0 E4: 8 E5: 21</td>
</tr>
</tbody>
</table>

*Note. One paper in Academic Search Premier and one in ERIC, are repeated in SpringerLink.*

From the selected paper, the following details were examined: the purpose of the study, the sample size of the students involved in the experiments, the description of the setting, data collection and analysis methods, presented results and the implication of the studies.

**Discussion**

Four major factors are focused on in this paper: the type of studies conducted, the robot use’s influence on child behaviour and development, stakeholder perception, and children’s reaction to robot design or appearance.

**Types of studies conducted**

Majority of the reviewed papers employed non-experimental studies. There were three studies involving the use of survey, where video was used to record children’s behaviour and interaction with the robots. Four quasi-experimental studies involved pre-test and post-test, which were conducted with control group. There were ten anecdotal case studies, five mixed-method studies and one correlational study. There were three experimental studies and one short review paper. The detail of each study approach is listed in Table 3.

Table 3. Types of study reported in the reviewed papers

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-experimental (Mixed-method Study)</td>
<td>Williams et al., 2007; Levy &amp; Mioduser, 2008; Liu, 2010; Young et al., 2010; Sugimoto, 2011</td>
</tr>
<tr>
<td>Non-experimental (Anecdotal case studies)</td>
<td>Barker &amp; Ansorge, 2007; Rusk et al., 2008; Highfield, 2010; Hong et al., 2011; Chang et al., 2010; Chen, Quadir &amp; Teng, 2011; Slangen et al., 2011; Varney et al., 2012</td>
</tr>
<tr>
<td>Non-experimental (Cross-sectional survey)</td>
<td>Woods, 2006; Lin et al., 2012</td>
</tr>
<tr>
<td>Non-experimental (Longitudinal survey study)</td>
<td>Ruiz-del-Solar &amp; Avilés, 2004</td>
</tr>
<tr>
<td>Non-experimental (Case studies)</td>
<td>Bers, 2010; Bers &amp; Portsmore, 2005</td>
</tr>
<tr>
<td>Non-experimental (Correlational study)</td>
<td>Bers, 2010</td>
</tr>
<tr>
<td>Quasi experimental (Pre-test &amp; Post-test)</td>
<td>Barker &amp; Ansorge, 2007; Whittier &amp; Robinson, 2007; Chambers et al., 2008; Kazakoff et al., 2013</td>
</tr>
<tr>
<td>Experiment study</td>
<td>Beran et al., 2011; Salter et al., 2004; Michaud et al., 2005</td>
</tr>
<tr>
<td>Short review paper</td>
<td>Cangelosi et al., 2010</td>
</tr>
</tbody>
</table>
Robot’s influence on children’s behaviour and development

The reviewed articles revealed four major themes where robot was able to aid in child’s behaviour or development.

Theme 1: Problem-solving abilities, team skills and collaboration

Studies by Barak (2009) and Varney et al. (2012) were conducted to investigate how the introduction of robots could change education, especially to help prepare children with 21st century skills and to increase student interest in robotics. The study conducted by Barak (2009) showed that high school students were able to come up with inventive solutions to problems and could benefit from working on project-based programmes. Robotic kits such as LEGO Mindstorm allowed students to work in teams as they carried out their projects in small groups.

Robotics was further viewed as an effective tool to develop “team skills” in students (Varney et al., 2012). The use of robots in various activities with young children supports constructivism as a learning method. Students discuss, solve problems, work with their peers, and combine their knowledge in order to construct their robots. In Chang et al. (2010), the results from the study further supported that robots could create an interactive and engaging learning experience.

Robots in elementary school helped promote collaboration and problem-solving skills in children as they became involved in the process and construction of their artefacts for their robotic projects. This was further highlighted by Hong et al. (2011) study where robots allowed children to engage in deep reflection as they solve problems and collaborate with their peers, both of which enhanced their learning experience.

Theme 2: Achievement scores, science concepts and sequencing skills

The study conducted by Baker and Ansorge (2007) examined students’ achievement scores with the use of robots in their science curriculum. Robots were found to be effective at teaching 9-11 year old students science, engineering and technical concepts. Results from another experiment study conducted by Kazakoff et al. (2013) supported the use of the robotic programming such as CHERP, a tangible programme which helped increase sequencing skills in pre-kindergarten and kindergarten children.

Table 4. Articles that reported on skills development

<table>
<thead>
<tr>
<th>Papers</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barker &amp; Ansorge, 2007</td>
<td>Results showed increase mean scores from pre- to post-test, indicating that robotics was effective at teaching youth about science, engineering, &amp; technology concepts.</td>
</tr>
<tr>
<td>Williams et al., 2007</td>
<td>Study shows a significant difference on acquiring physics knowledge but not for science inquiry skills</td>
</tr>
<tr>
<td>Barak, 2009</td>
<td>Study reveals that students often come up with inventive solutions to problem when learning with robots.</td>
</tr>
<tr>
<td>Highfield, 2010</td>
<td>The result significantly showed that children engaged in multiple mathematical processes; they demonstrated perseverance, motivation &amp; responsiveness.</td>
</tr>
<tr>
<td>Whittier &amp; Robinson, 2007</td>
<td>The results showed that all students obtained significant gains in their conceptual understanding. There is an increase of mean pre-test from 26.9% to post-test 42.3%.</td>
</tr>
<tr>
<td>Kazakoff et al., 2013</td>
<td>Results indicated that the sequencing ability of pre-kindergarten and kindergarten students increases when participating in an intensive robotics and programming curriculum.</td>
</tr>
<tr>
<td>Bers, 2010</td>
<td>The result showed that boys had a higher mean score than girls on more than half of the tasks. Boys scored significantly higher than girls in properly attaching robotic components and programming using ‘Ifs’.</td>
</tr>
<tr>
<td>Slangen et al., 2011</td>
<td>Robots helped challenge pupils to manipulate, reason, predict, hypothesize, analyze and test.</td>
</tr>
</tbody>
</table>
The use of robot to assist non-English speaking students to improve in their understanding of science concepts was carried out by the Whittier and Robinson (2007) study. Results showed that all students obtained sufficient gains in their science conceptual knowledge with an increase from 26.9% in pre-test to 42.3% in post-test. The middle school students developed problem-solving skills, inquiry and engineering design skills. Robots were also used to develop and improve learning of science concepts, technology and problem-solving, which was further supported by Barak’s (2009) qualitative analysis of observations, interviews and reflections of students working on their projects. Similarly, anecdotal records in the Highfield (2010) study showed that robotic toys could be catalyst for mathematical problem solving through participation in multi-faceted approach by integrating and inter-relating concepts and skills through dynamic tasks. The use of robotic to develop of physics content knowledge showed a significant difference but not for the science inquiry skills, according to the Williams et al. (2007) study. Table 4 shows a summary of the skills where robot has a positive effect.

**Theme 3: Language skills development**

In the study by Chang et al. (2010), a humanoid robot was used to teach a second language in a primary school. Results showed that robots could create interactive and engaging learning experiences as the children responded with high motivation. The use of robots for language development was found to be advantageous as it also allowed for demonstration of highly mobile behaviour and extensive repetition. Sugimoto (2011) used robot for storytelling, where the robot was used in students’ learning and provided opportunity for children to learn in a mixed-reality environment. The children engaged strongly in story expression and acted in a coordinated manner while also being involved in their story creation with their robots.

<table>
<thead>
<tr>
<th>Papers</th>
<th>Overview of paper on language skills development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang et al., 2010</td>
<td>Results indicate that robots could create interactive and engaging learning experience for students.</td>
</tr>
<tr>
<td>Young et al., 2010</td>
<td>Quantitative results showed that 95% have positive attitude towards tangible learning companions/robots. They become more active in practicing conversation.</td>
</tr>
<tr>
<td>Hong et al., 2011</td>
<td>Students were highly involved and reflective during the construction of their artefacts.</td>
</tr>
<tr>
<td>Varney et al., 2012</td>
<td>Results showed that robot could be used as an effective tool in children to develop ‘team skills’; 75% of students actively raised questions.</td>
</tr>
<tr>
<td>Sugimoto, 2011</td>
<td>In the study, the children engage strongly in story expression and acted in a coordinated manner.</td>
</tr>
<tr>
<td>Chambers et al., 2008</td>
<td>Results suggested that providing children with physical experiences were not sufficient to understand mechanical concepts. Timely &amp; appropriate intervention is important.</td>
</tr>
<tr>
<td>Bers, 2010</td>
<td>TangibleK robotics could be implemented in the early childhood setting in a developmentally appropriate way by integrating other disciplines.</td>
</tr>
<tr>
<td>Rusk et al., 2008</td>
<td>Results suggested multiple paths for engagement of children, teens, families and educators.</td>
</tr>
<tr>
<td>Levy &amp; Mioduser, 2008</td>
<td>The role of adult’s interaction enables children to shift into more complex technological rules.</td>
</tr>
<tr>
<td>Varney et al., 2012</td>
<td>The study presented results on the efficacy of the LEGO robotic programme in fostering student’s interest.</td>
</tr>
<tr>
<td>Ruiz-del-Solar &amp; Avilés, 2004</td>
<td>Social robots were effective in fostering students’ interest in engineering.</td>
</tr>
<tr>
<td>Michaud, et al., 2005</td>
<td>Roball, a robot capable of autonomous motion, was used in child-development studies.</td>
</tr>
<tr>
<td>Cangelosi et al., 2010</td>
<td>Studied embodied cognitive agent-humoid robot. Discussed areas such as complex sensorimotor, linguistic &amp; social learning skills.</td>
</tr>
<tr>
<td>Chen, Quadir &amp; Teng, 2011</td>
<td>The use of robot with computer and book enhanced students’ concentration in their learning of English, interest and motivation.</td>
</tr>
</tbody>
</table>
According to Slangen et al. (2011), students working on projects using LEGO and Mindstorms were found to be involved in frequent process of comparing their test results with their objectives, expectations, and in refining their conceptual knowledge and skills. Table 5 summarizes the articles that reported on the use of robots for language skills development.

**Theme 4: Participation**

Rusk et al. (2008) introduced Picocritter invention kit program to increase participation from children, teens, families and educators in robotics-related endeavors via workshops, after-school programs and professional development programs. The workshops allowed students to work on broad themes based on their own interests. As these students were given the opportunity to combine art and engineering, encouraged to use storytelling and exhibition and introduced to new technologies, their interest in robotics increased.

**Parents’, educators’, and children’s perception of educational robots**

Liu (2012) and Ruiz-del-Solar and Avilés (2004) investigated perception of parents, children and teachers on the use of educational robots. The results from Lin et al. (2012) revealed that most parents’ would consider educational robots as beneficial for their children. However, parents felt that they were less confident when playing and teaching their children on using robots.

Ruiz-del-Solar and Avilés (2004) studied the children’s degree of satisfaction on robot use, their inquired level of competence and their eventual interest to pursue an engineering career. 700 children and teachers were surveyed in that study and 86% of the participants would consider studying in an engineering or science university in the future.

In the Bers (2010) study, educators developed computational thinking and learning about the engineering design process in young children by introducing the TangibleK programme. It integrated other disciplinary learning in a developmentally appropriate way for young children. Table 6 provides the list of articles and their reports on stakeholder perception on using robots in education.

<table>
<thead>
<tr>
<th>Papers</th>
<th>Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beran et al., 2011</td>
<td>Results from frequency and content analysis suggested that a significant proportion of children ascribe cognitive, behavioural, and affective characteristics to robots.</td>
</tr>
<tr>
<td>Salter et al., 2004</td>
<td>Findings suggested that touch could have an important role to play when developing natural human-robot interfaces. It further suggested that robot interaction levels could vary to suit different children.</td>
</tr>
<tr>
<td>Woods, 2006</td>
<td>Results showed that although the robots are very human-like, children were still capable of distinguishing them from humans. However, the robots evoke a feeling of discomfort or repulsion.</td>
</tr>
<tr>
<td>Liu, 2010</td>
<td>Results showed children regard</td>
</tr>
<tr>
<td></td>
<td>- Educational robot as a plaything;</td>
</tr>
<tr>
<td></td>
<td>- Studying robotics as a source of employment;</td>
</tr>
<tr>
<td></td>
<td>- Learning of robotics as a way to high tech.</td>
</tr>
<tr>
<td></td>
<td>Male and female perceptions differ.</td>
</tr>
<tr>
<td>Lin et al., 2012</td>
<td>Results indicated that parents considered educational robots as beneficial for their children. But they were less confidence in playing and teaching with educational robots with their children themselves.</td>
</tr>
<tr>
<td>Bers &amp; Portsmore, 2005</td>
<td>Engineering students gained insight into the educational system and issues involved in incorporating ICT into the classroom. Pre-service teachers saw the potential offered by technology and what they would need to know to continue using it.</td>
</tr>
</tbody>
</table>
Children’s reaction to robot’s design or appearance

Levy and Mioduser (2008) presented rich anecdotal data on children’s descriptions and explanations of robots’ behaviour. Their study involved children in two strands of tasks (description and construction). It also showed that when adult facilitate and interact with the children, they were capable of shifting into more complex technological rules. In addition, a study conducted with 184 (Beran et al., 2011) showed that a significant proportion of the children ascribe cognitive, behavioural, and affective characteristics to robots.

159 children were asked to evaluate 40 images of robot through questionnaires in order to investigate how children perceive robot’s appearance (Woods, 2006). The study showed that children perceive robots’ intentions and capabilities based on robot appearance. Children judged human-like robots as aggressive and machine-like ones as friendly. Sullivan and Bers (2012) showed using the TangibleK programme that the boys scored significantly higher than girls in properly attaching robot components and programming using “Ifs.” However, as reported for the rest of the tasks gender differences were statistically insignificant.

Conclusion

The effectiveness of robots in education programme could be analyzed from different aspects: Study design in order to report meaningful and statistically significant results, robot’s effects on child’s behaviour and development, relevance of stakeholders’ perception on using robots in and outside of classroom setting, and users’ reaction (especially the children) to the robot’s design.

Researchers, majority of whom relied on non-experimental methods, implemented various approach to validate their studies. However this just shows that experimental methods are sorely lacking; quantitative analysis is needed, as pointed by Benitti (2012).

In education, the use of robots has the potential to help children develop various academic skills like science process understanding, mathematical concept development and improvement of achievement scores (Barker & Ansorge, 2007; Williams et al., 2007; Highfield, 2010). In addition, the introduction of robotics in curriculum also increases children’s interest in engineering. As reported in Chang et al., 2010, the use of robots in education allows children to engage in interactive and engaging learning experiences. Robots seem appropriate to use in language skill development because it allow for a richer interaction (Sugimoto, 2011; Chambers et al., 2008; Bers, 2010; Chang et al., 2010; Young et al., 2010).

Two new factors have emerged in this review paper: the stakeholder’s perception and the value of robot design. Aside from the main users (children), parents and educators have to be on-board as well in order to increase the chances of success of this kind of programmes. Lack of parental support would confine educational robots to applications only inside the classroom.

Lastly, design is usually the last consideration when incorporating robots into an application. However, as Woods (2006) and Sullivan & Bers (2013) studies showed, design could make a difference on robot perception and hence, how the children would interact with it. Unfortunately, not a lot of work has been done yet on this question.

Past studies are like beacons on where research have been and indicates various milestones (e.g., Cangelosi et al., 2010). This paper shows a possible roadmap and highlights research gaps in this field.

Acknowledgements

The authors would like to thank the Singapore Millennium Foundation for supporting this work.
References


## Appendix

### Appendix Table. Details of the selected studies

<table>
<thead>
<tr>
<th>Paper</th>
<th>Level (age)</th>
<th>Area explored</th>
<th>Robot used</th>
<th>Study detail</th>
<th>Results</th>
<th>Implications</th>
<th>Type of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barker &amp; Ansorge, 2007</td>
<td>32 (9-11 years old)</td>
<td>Achievement Scores</td>
<td>LEGO Mindstorms</td>
<td>28 lessons conducted using experiential learning modes to teach Science Engineering Technical concepts</td>
<td>No significant results between Pre-test &amp; Post-test for control group. Robot group showed a significant increase from ( M = 7.93, SD = 3.71 ) to ( M = 17.00, SD = 0.88 )</td>
<td>Increase of mean scores from pre-test to post-test indicated that Robotic group was effective at teaching youth about SET concepts.</td>
<td>Quasi experimental study</td>
</tr>
<tr>
<td>Williams et al., 2007</td>
<td>K-12 (21 middle school students)</td>
<td>Acquisition of Physics Content Knowledge and Scientific Inquiry Skills</td>
<td>LEGO Mindstorms and ROBOLAB</td>
<td>2 weeks robotic camp as students work in small groups to examine whether they increase their Physics Content Knowledge &amp; Science Inquiry Skills</td>
<td>There is a significant difference on the Physics Content Knowledge but not for the Science inquiry skills.</td>
<td>Mixed methods</td>
<td></td>
</tr>
<tr>
<td>Rusk et al., 2008</td>
<td>Robotic activities were arranged for * museum workshop for families * after-school program for girls * professional-development workshop for educators.</td>
<td>Broadening of participation in robotic</td>
<td>Picocricket</td>
<td>Robotic workshop for students to work on themes to foster their interest and a sense of shared experiences. Combining art and engineering encourage story-telling, exhibition &amp; new technologies.</td>
<td>The results suggested multiple paths for engagement for children, teens, families, and educators.</td>
<td>Robotic is introduced in areas of students’ interest e.g., music, art and story-telling, providing new learning experiences to wider audience.</td>
<td>Non-experimental (Anecdotal Records, Case Studies)</td>
</tr>
<tr>
<td>Levy &amp; Mioduser, 2008</td>
<td>Kindergarten 3 boys, 3 girls, randomly selected, (5yrs - 6yrs old)</td>
<td>Children’s perspectives</td>
<td>LEGO mobile robots</td>
<td>To investigate children’s perspectives. Children took part in a sequence braided of two strands of tasks: Description and Construction. Five 30-45 minute session. Data collected on children’s description and explanations of robots’ behaviour.</td>
<td>The role of adult during facilitation: with adult’s interaction, children shift into more complex technological rules.</td>
<td>Learning is viewed as enculturation and knowledge is socially constructed. Differentiate between technological and psychological points of view.</td>
<td>Mixed-method</td>
</tr>
<tr>
<td>Barak, 2009</td>
<td>Junior High School, 80 students</td>
<td>To improve learning concepts in Science, Technology and problem-solving</td>
<td>LEGO Mindstorms</td>
<td>Data are collected through qualitative analysis of observations, interviews &amp;</td>
<td>Students often come up with inventive solutions to problems. They are likely to</td>
<td>Non-experimental (Anecdotal Records, Case studies)</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Grade</td>
<td>Sample Size</td>
<td>Methodology</td>
<td>Math Tool</td>
<td>Study Design</td>
<td>Results</td>
<td>Differences</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----------</td>
<td>--------------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>Liu, 2010 Elementary: Grades 4-6</td>
<td></td>
<td></td>
<td>Mixed-method</td>
<td>LEGO Mindstorms &amp; robots</td>
<td>Questionnaire</td>
<td>Results showed: (1) children regard educational robot as a plaything, (2) learning about robots as a source of employment, (3) learning of robotics as a way to high technology; Differences between male and female perception.</td>
<td></td>
</tr>
<tr>
<td>Highfield, 2010 33 (3-4 years old) 22 (Year 1)</td>
<td></td>
<td></td>
<td>Anecdotal, Case Studies</td>
<td>Bee-bots &amp; Pro-bots</td>
<td>Task design</td>
<td>The result showed significant children engagement in multiple mathematical processes; they demonstrated perseverance, motivation &amp; responsiveness.</td>
<td></td>
</tr>
<tr>
<td>Chen, Quadir &amp; Teng (2011) Elementary School, 5 students</td>
<td></td>
<td></td>
<td>Anecdotal Case studies</td>
<td>Robot, Zigbee, computer and book</td>
<td>Observation and interview</td>
<td>Use of computer with robot and book provided interactive experiences to students</td>
<td></td>
</tr>
<tr>
<td>Young et al., 2010 Elementary School 68 (Grade 3-4); 6 students (2 boys &amp; 4 girls) were selected as a focus group</td>
<td></td>
<td></td>
<td>Mixed-method</td>
<td>Rocky robot</td>
<td>Questionnaire survey conducted in an elementary school in Taiwan</td>
<td>The children were active in practising conversation with the robot</td>
<td></td>
</tr>
<tr>
<td>Hong et al., 2011 Elementary School Collaboration of learning in technological project design</td>
<td></td>
<td></td>
<td>Non-experimental (Anecdotal records)</td>
<td>POWERTECH robot</td>
<td>Students took part in a POWERTECH contest in Taiwan</td>
<td>Reflection essential for problem-solving were often raised</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Grade</td>
<td>School</td>
<td>Participants</td>
<td>Intervention</td>
<td>Observation</td>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Lin et al., 2012</td>
<td>Junior High School's parents: 39; 17 male, 22 female</td>
<td>Parents’ perceptions towards educational robots</td>
<td>Questionnaire survey about the parent’s attitude</td>
<td>Results indicated that parents considered educational robots as beneficial for their children.</td>
<td>Parents were willing to provide chance and encourage children to learn with educational robots.</td>
<td>Non-experimental (Cross-sectional-survey study)</td>
<td></td>
</tr>
<tr>
<td>Ruiz-del-Solar &amp; Avilés, 2004</td>
<td>K-12 700 children and teachers in Chile</td>
<td>Children and teachers’ perception of educational robots</td>
<td>Reviews on use of robots since 2000 through surveys with children and teachers.</td>
<td>Children and teachers’ perception of educational robots.</td>
<td>92% satisfied with the workshop, 88% finished all the basic tasks during the workshop, 86% indicated they would follow an engineering or science career in the future.</td>
<td>Non-experimental (Longitudinal Study-using survey)</td>
<td></td>
</tr>
<tr>
<td>Varney et al., 2012</td>
<td>Elementary school</td>
<td>TASEM summer camp to raise interest in STEM</td>
<td>Working in small groups, 1 hr/week session</td>
<td>The robotic programme allowed students of different socio-economic and cultural backgrounds to participate.</td>
<td>The children engage strongly in story expression processes and acted in a coordinated manner.</td>
<td>Mixed-method</td>
<td></td>
</tr>
<tr>
<td>Sugimoto, 2011</td>
<td>Elementary school</td>
<td>A mobile mixed-reality environment. Study conducted over 2 weekends.</td>
<td>Children took part in a story creation by manipulating a robot and a handheld projector. The study involved 2 previous pilot studies.</td>
<td>The children engaged strongly in story expression processes and acted in a coordinated manner.</td>
<td>Anecdotal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Grade Level</td>
<td>Group</td>
<td>Instructional Tool</td>
<td>Methodology</td>
<td>Results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>-------</td>
<td>-------------------</td>
<td>-------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chambers et al., 2008</td>
<td>Elementary</td>
<td>9-10 yrs</td>
<td>Creative Product Semantic Scale</td>
<td>Quasi-experimental</td>
<td>Quantitative results were collected with the use of Creative Product Semantic Scale on their story creation. Results suggest that providing children with physical experiences were not sufficient to develop mechanical conceptual understanding, of the importance of timely and appropriate intervention. Results confirm that there is variability among children in how they reason about gears &amp; conceptual development. A guided inquiry instructional approach is proposed for the conceptual understanding development.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chang et al., 2010</td>
<td>5th graders</td>
<td></td>
<td>Instructional tool for 2nd language</td>
<td>Non-experimental (Case studies-observational records)</td>
<td>(1) The humanoid robot performs rich gestures. Non-verbal signals are important part of communication. (2) The robot can change intonation or speech rate. (3) The human appearance of a robot attracted attention, even from weaker students. This may motivate them to participate more in the language class. (4) Robots’ ability to interact and recognize students’ commands offer a more natural way to perform.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author(s)</td>
<td>Grade or Age</td>
<td>Description</td>
<td>Assessment</td>
<td>Development of Evidence</td>
<td>Research Design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bers, 2010</td>
<td>Prekindergarten to 2nd grade</td>
<td>To develop computational thinking &amp; learning about the engineering design process in young children</td>
<td>Tangible K-programme language drills.</td>
<td>TangibleK robotics was implemented in the early childhood setting by integrating it with other disciplinary learning in a developmentally appropriate way for young children.</td>
<td>Non-experimental (Case studies-observational records)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whittier &amp; Robinson, 2007</td>
<td>Middle school (Grade 7-8), 29 students (16 Grade 7, 13 Grade 8)</td>
<td>Using robotics to teach non-English proficient students in developing their understanding of science concepts</td>
<td>LEGO, Evobots 12 sessions of 60-minute lessons. Teachers use LEGO robotics to address state science standards.</td>
<td>The results showed that students having significant gains in their conceptual understanding. An increase of mean pretest 26.9% to posttest 42.3%. Results from frequency and content analysis suggest that a significant proportion of children ascribe cognitive, behavioural, especially affective characteristics to robots.</td>
<td>Quasi Experiment study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beran et al., 2011</td>
<td>184 children, 5-16 years old, 98 female, 86 male</td>
<td>Using robotics to teach non-English proficient students in developing their understanding of science concepts</td>
<td>A 5 degree freedom robot arm, performing block stacking task Semi-structured interviews conducted with the children. 9 questions were asked whether the robots referenced humanistic qualities.</td>
<td>Students developed many science processes, problems-solving, inquiry, and engineering design skills. Results from frequency and content analysis suggest that a significant proportion of children ascribe cognitive, behavioural, especially affective characteristics to robots.</td>
<td>Experiment Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salter et al., 2004</td>
<td>6 Children (5-7 years old)</td>
<td>Autora project develop for use with children with autism in therapeutic and educational context.</td>
<td>Pekee robot Children grouped into clusters according to their psychological classification. Sensor captures children’s interaction level.</td>
<td>Results indicated that robot’s behaviour can be adapted to a different children. It is suggested for future to use robot to quantify and assess children’s behaviour.</td>
<td>Experiment study using sensor and observational techniques</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woods, 2006</td>
<td>159 children</td>
<td>To examine children’s perception of robots’ appearance</td>
<td>Evaluate 40 robot images by completing a questionnaire on appearance, personality and emotions</td>
<td>Some robots are human-like but still distinguishable from humans and evoke a feeling of discomfort or</td>
<td>Non-experimental (Cross-sectional survey)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
understanding capabilities and emotional expression.

Children judged human-like robots as aggressive, but human–machine robots as friendly.

A paired t-test was conducted on the children sequencing abilities using sequencing cards. A pre-test and post-test conducted. There was a control group.

Results indicated that it was possible to see increases in the sequencing ability of pre-kindergarten and kindergarten students participating in a robotics and programming curriculum in as little as 1 week.

Botrics offer children and teachers a new way to tangibly interact with traditional early childhood curricular themes.

Pre-service teachers working in partnership with engineering students during their training. The goal is to develop a model and approach for this teaching methodology.

Three models were evaluated:
- Developer’s Model
- External Consultant’s Mode
- Collaborator’s Model

From engineering’s perspective, students gained insight into the educational system and issues involved in incorporating ICT into the classroom.

Pre-service teachers saw the potential of the technology and resources needed to continue using it.

A study on gender differences in robotics and programming achievement

The study examined whether kindergarten boys and girls were equally successful in a series of building and programming tasks. The TangibleK Program consisted of a six robotics lessons.

Results showed that boys had a higher mean score than girls on more than half of the tasks but very few differences in the results were statistically significant.

Boys scored significantly higher than girls in only 2 areas: properly attaching

162
<table>
<thead>
<tr>
<th>Study</th>
<th>Age Range</th>
<th>Methodology</th>
<th>Robot/Tool</th>
<th>Description</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slangen et al., 2011</td>
<td>10-12 year olds</td>
<td>10-12 year olds</td>
<td>LEGO Mindstorms NXT</td>
<td>Study on conceptual and cognitive analysis to develop a reference frame to determine students’ understanding of robotics</td>
<td>Study concluded that robotic DMEs challenge pupils to manipulate, reason, predict, hypothesize, analyze and test. Students frequently compare test results with their objectives and expectations to refine their conceptual knowledge and skills.</td>
</tr>
<tr>
<td>Michaud, et al., 2005</td>
<td>12-24 months children</td>
<td>Non-experimental (anecdotal study)</td>
<td>Roball</td>
<td>To examine the potential of using robot to help children in areas of their language, affects, motor, intellectual &amp; social skills development</td>
<td>Roball could capture children’s attention, enabling interaction studies.</td>
</tr>
<tr>
<td>Cangelosi et al., 2010</td>
<td>Humanoid robot</td>
<td>Experimental</td>
<td>Humanoid robot</td>
<td>Study of embodied cognitive agents to understand cognitive development, complex sensorimotor, linguistic and social learning skills</td>
<td>The milestones on the roadmap directs future work of cognitive development.</td>
</tr>
</tbody>
</table>

*Note: Legend for study types: 6pt Likert-scale assessment tool; Non-experimental (anecdotal study)
ABSTRACT

This article describes an empirical investigation of technology-enhanced peacemaking in a conflict-stressed school environment. The peacemaking intervention required students in conflict-laden groups to collaborate on various game-like learning activities on a multitouch interactive tabletop, over the span of three weeks. Student interviews and video observations provided evidence that tabletops can become a means for communication and collaboration giving the chance for students in conflict to share a common space, shifting attitudes and improving their relationships. The study elaborates on the affordances of tabletops as they become apparent in the context of peacemaking, unpacking the still widely unexplored potential of multitouch interactive technology in peace education.

Keywords

Technology-enhanced peacemaking, Technology-enhanced learning, Peace education, School conflict, Technology for social change, Technology for peace, Interactive tables, Tabletops, Collocated collaboration, CSCL, CSCW

Introduction

Promoting peace or educating for peace in school contexts through the means of technology is a topic we rarely see addressed. Although technology has been extensively employed to improve instructional design, development and delivery of instructional content, and access to information, its potential to foster peace within educational interventions is relatively unexplored (Veletsianos & Eliadou, 2009; Buckner & Kim, 2011).

Veletsianos and Eliadou (2009) conducted a systematic review of research on the use of technology in educational initiatives to promote peace and peace-related outcomes. Their exhaustive search resulted in only 37 relevant works in the period 1990-2009 out of which only five presented empirical evaluations of technology-infused interventions aimed at promoting peace outcomes (the remaining included theoretical papers or descriptions of technology-infused interventions without empirical data). The focus of these 37 efforts, revolved around the use of the Internet technologies mediating long-distance collaboration, in particular: (i) virtual learning environments to host learning activities related to peace, (ii) Web 2.0 technologies (blogs, wikis, social networks and video sharing sites), Email, and video conferencing to promote communication and collaboration among people of diverse backgrounds, and (iii) serious games or MMORPGs to promote exposure to diverse populations and social interaction, such as the PeaceMaker game (Burak, Keylor, & Sweeney, 2005) which simulates the Israeli-Palestinian conflict and engages dyads in negotiating peace (Veletsianos & Eliadou, 2009).

Building on Veletsianos and Eliadou (2009) and further reviewing literature from 2009 to 2015, we argue that empirical work on technology-enhanced peace interventions in educational contexts continues to be very limited and sporadic. Internet-based interventions dominate any efforts, for example using mobile devices to share narratives of digital stories with a broader community (Buckner & Kim, 2011) or learning about conflict management through simulations and games (Brynen & Milante, 2013; Gehlbach et al., 2008). Collocated technology-enhanced peacemaking educational interventions are virtually non-existent in the literature.

We set the present work on the basis that schools can be conflict-stressed environments where peacemaking actions are needed, yet very little work in this area has taken advantage of the current educational innovations. We seek to investigate the mediating and supportive role of multitouch interactive tables (tabletops) in shaping relationships and reducing conflict among students, recognizing the potential of this technology to redefine collocated collaboration and communication. We further consider physicality and collocation to be important factors in peacemaking efforts, especially in cases where language barriers to effective communication exist. The study took place in a conflict-stressed school environment where increased incidents of violence and delinquency existed (including verbal and
physical conflict and bullying). The leading research question of the study was: Do tabletops provide a means for communication and collaboration, enabling students in conflict to share a common space, perhaps breaking down barriers? Given the lack of research in the area, we consider this work timely and relevant, while serving an important purpose.

**Mutlitouch interactive tables**

Tabletops are large horizontal displays that enable interaction by multiple concurrent users. They are relatively new technology, being commercially available only since 2009 with the introduction of MS Surface. Despite this, there is substantial research around them. Tabletops have attracted the attention of designers of formal (Dillenbourg & Evans, 2011; Higgins, Mercier, Burd & Hatch, 2011) and informal (Davis et al., 2015) learning experiences as highly supportive systems for collaboration and interaction.

Briefly, investigations of tabletops have shown that being able to see another’s physical actions on the shared display can enhance awareness, which in turn can support fluid interaction and coordination (Hornecker, Marshall, Dalton, & Rogers, 2008; Fleck et al., 2009). Also, their horizontal orientation allows users to hover their hands easily over the surface, and as a result, gesture-based communication can supplement or even replace verbal communication (Rick, Marshall, & Yuill, 2011). Tabletops have been found to encourage playfulness in the interaction. For example, Piper, Friedman, & Hollan (2012) discussed how the touch-input allowed freedom and playfulness in students’ interactions which differed from paper/pen-based interactions. Also, Jamil, O'Hara, Perry, Karnik, and Subramanian (2011) demonstrated how specific interaction techniques promoted playfulness in students’ interactions during a learning task. Moreover, tabletops have been found to enhance the sense of teamwork (Morris, Huang, Paepcke, & Winograd, 2006), “invite” interaction and willingness to participate in groups tasks (Rogers & Lindley, 2004), increase equity in physical interaction (e.g., Ioannou, Zaphiris, Loizides, & Vasiliou, 2013b; Marshall et al., 2008), promote joint attention on the task (Fleck et al., 2009; Higgins et al., 2011), and improve the (learning) experience and engagement with the task (Buisine et al., 2012; Ioannou, Christofi, & Vasiliou, 2013a; Ioannou, Zenios, & Stylianou, 2014).

**Tabletops and peacemaking**

Despite the substantial work on the use of tabletops in formal and informal learning settings, their affordances for peacemaking in school contexts have yet to be explored. A couple of previous works have considered tabletops as “peace technology” but none of them was conducted in a natural school environment with young students in conflict.

First, Stock et al. (2008) designed a collocated tabletop interface as a tool for reconciliation. The so called NNR-Table allowed multimedia narrations to be contributed from two opposing sides; participants worked together to achieve a narration acceptable to both viewpoints (i.e., by revising and completing the narration together). Interventions with the use of NNR-Table were found successful in helping Jewish-Arab pairs of youth reach a compromise and learn more about each-other’s viewpoints (Stock et al., 2008; Zancanaro et al., 2012). Second, Ioannou, Zaphiris, Loizides, and Vasiliou (2013b) designed a collocated brainstorming tabletop activity to facilitate dialog and consensus decision-making in groups of college-students discussing sensitive and controversial topics, including peace-building in a country of long-term ethnic conflict. The authors found that discussion around the tabletop was fluent with no evidence of tension, anxiety or strong disagreement among the participants. Furthermore, although not in the context of peace, other studies have presented affordances of tabletops that are relevant and pertinent to this work. For example, it is argued that, because tabletops support concurrent input and shared control, they enable collaborators to negotiate conflict (Fleck et al., 2009; Falcão, & Price, 2011) and can help engage users in “creative conflict” that is, arguing and disagreeing directed at ideas rather than people (Basher, Burd, & Baghaei, 2013).

**Method**

**Participants and setting**

The participants were twenty (20) students 9-11 years old at a small (80 students) public elementary school in the eastern Mediterranean. The school is located in an area with low socio-economic indicators and is characterized by
large numbers of foreign students (i.e., minority enrollment), high drop-out rate and increased incidences of conflict (verbal and physical), violence and delinquency. The school belongs to the “Zone of Educational Priority” – an educational initiative linked to Europe’s tactic to support students from areas with low social and economic indicators in order to reduce dropout and promote educational success. All students attend (official) language lessons as the National Curriculum requires. Yet, many of them have limited oral and written communication abilities, perhaps because in many cases a different language is spoken at home. To mitigate the frequent incidences of violence and delinquency, teachers in this school make efforts to infuse conflict management principles into the everyday curriculum helping students to understand conflict and handle anger appropriately. Also, the National curricula on Health and Life Education covers topics such as the development of personal, social and communication skills, effective management of emotions, promotion of self-esteem and improving the psychosocial school climate. Despite these efforts, the school continues to be a conflict-stressed environment.

We collaborated with the school teachers of grades 4-6 to decide (i) the synthesis of the groups and (ii) the peacemaking tabletop activities. These teachers were 3-5 years in this school and were familiar with ongoing conflict between particular students (i.e., physical and verbal conflict and bullying) beyond temporal or typical fights and arguments. Based on this knowledge, five groups of four students were nominated to participate in the study: two groups of 4th-graders, two groups of 5th-graders, and one group of 6th-graders. The teachers’ perceptions of conflict between students were cross-checked with students’ self-reports during pre-intervention interviews. In fact, the proposed grouping for the intervention was finalized after student’ self-reports of regular and ongoing conflicts with specific classmates, which was fully consistent with the teachers’ input. Within-group conflict had various forms, e.g., all four students in conflict between them, or two pairs in conflict, or three students in conflict with the 4th group-member. Table 1 presents the synthesis of the participating groups in terms of gender split, ethnic minority and within-group conflict.

<table>
<thead>
<tr>
<th>Group name</th>
<th>Grade-level</th>
<th>Gender split</th>
<th>Ethnic minority group-members</th>
<th>Group-members in conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>4th-grade</td>
<td>3 girls/1 boy</td>
<td>3 MO, KA, AI in conflict with AL</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>5th-grade</td>
<td>2 girls/2 boys</td>
<td>2 AA, CH, DI in conflict with AS; CH in conflict with DI</td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>5th-grade</td>
<td>2 girls/2 boys</td>
<td>4 RO in conflict with ST, AN, OS; OS in conflict with RO, AN, ST</td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>4th-grade</td>
<td>4 girls</td>
<td>4 MA in conflict with AS; KL in conflict with AP</td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>6th-grade</td>
<td>1 girl/3 boys</td>
<td>4 AX, XO, RU, OM all in conflict between them</td>
<td></td>
</tr>
</tbody>
</table>

**Peacemaking activities**

Building on the idea of gamifying the learning experience (Garris, Ahlers, & Driskell, 2002), we aimed to create an engaging and fun learning environment using applications in which the educational content was combined with game characteristics. Additionally, we aimed to engage students in tasks which did not require extensive language use, considering the participants’ literacy levels as a potential barrier to participation; the use of a tabletop was deemed ideal in this case, affording gesture-based around verbal communication (Rick et al., 2011). Last, we sought “neutral” activities to turn attention away from peacemaking and life education as formal curricula activities. That is, we sought an intervention to bring students closer through play and teamwork, expanding on the English, math and the general education curricula.

We used the Samsung SUR40 tabletop which can respond to approximately 50 simultaneous inputs. We used a total of six applications. Four applications (designed for children aged 6-12) were freely downloaded from the Windows Apps Store - Education and Entertainment category (Age appropriate and free or very cheap tabletop applications can be found on the Windows Apps Store in all subject areas and can be easily customized and linked to aspects of the general education curricula.). Together with the school teachers we set the difficulty level of each application for
each participating group. Additionally, we used two custom-built applications developed at the [Lab Name, authors’ reference], populated with age-appropriate content, again in collaboration with the school teachers. In particular,

- iMath allowed students to practice addition, subtraction, multiplication, division, decimals and fractions through a series of calculation activities and games such as Math for Speed, Sudoku, and Match Card (Figure 1A).
- English Club allowed students to work on English (as a second language) vocabulary and spelling tasks through animated videos and games like Hangman and Wordsearch (Figure 1B).
- PuzzleTouch allowed students to entertain themselves with jigsaw puzzles of various difficulty levels (Figure 1C).
- Kids Play & Learn allowed students to interact with mini educational games such as learning the time, color mixing, and shapes (Figure 1D).
- IdeasMapping supports the collaborative classification of ideas or pictures (authors’ reference). We used this app with pictures related to recycling, four seasons, musical instruments and animal groups (Figure 1E).
- IdeaSpace supports collaborative story design and storytelling. This activity was based on a rich collection of images taken from a textbook specialized in civic education to present bullying at school, classroom conflict and family violence. Students were asked to collaborate on designing a story, which was saved and expanded in follow-up sessions (Figure 1F). Teacher facilitation was provided for this activity as storytelling was challenging for the participating students due to literacy levels.

The peacemaking activities run for three weeks in two 30-minutes intervention-sessions per week for each group, for a total of 180 minutes (or 3 hours) tabletop interaction per group for the duration of the study. Groups worked with two or three different tabletop applications per session (10-12 minutes per app) while the difficulty level was adjusted from session to session to maintain engagement considering carry-over effects (e.g., learning, fatigue). Students completed all activities associated with this investigation during school time.

**Data collection**

We conducted semi-structured interviews with the participating students at three time points: before the intervention (pre-interviews, 10-15 minutes each), at the completion of the intervention (post-interviews, 15-20 minutes each), and three months later (delayed-interviews, 10-15 minutes each). All interviews were undertaken by one of the school teachers who made students feel comfortable in talking about their feelings. At the time of pre-interviews, information about the upcoming intervention and the participating groups was not disclosed. The interviewer (knowing the potential synthesis of the groups) carefully extracted information about the relationships between teammates in a natural discussion about student’s daily school experiences and interactions with all their classmates, in and outside the classroom. Also, during the discussion, the interviewer asked the participants to rate their feelings about several classmates (potential teammates and others) on a Likert-type response scale from 1: Hate to 10: Love (scale with both numbers and smiley faces for better understanding). Last, a close-ended question concerning students’ access to technology was asked. The same questions regarding students’ relationships were repeated in the post and delayed interviews; a couple of additional questions were asked about students’ experiences working on the tabletop (in the post-interviews) and their memories of the experience (in the delayed-interview). During the intervention we video recorded students’ interactions around the tabletop for a total of 18 hours of video (3 hours per school visit). Two cameras were placed in the room for a full coverage of students’ body language (including gestures, posture and movement), facial clues, and tabletop activity together with their verbal communication.

**Results**

Two investigators worked closely together to examine the data corpus; consensus decisions were made on the observed patterns, while disagreements were resolved via discussion. First, the pre, post, and delayed interview data were fully transcribed by the interviewer, who was capable of understanding the participants’ talk and therefore, an accurate transcription of the dialog. The datasets were then examined (i) as separate corpuses and (ii) chronologically - tracking students’ talk about their teammates over time. The video data were split into episodes; an episode was typically 10-12 minutes long and involved a group’s work with one tabletop application. We looked at the episodes of the groups working on each application over time (e.g., group A working on iMath in sessions 1, 2, and 3).
Pre-interview data

Following Attride-Stirling’s (2001) thematic networks approach to analysis of textual data, we mapped the nature of conflict at school and the kinds of conflicts within the participating groups, based on the pre-interview data. Two organizing themes of conflict emerged – conflict during recess and conflict in the classroom – broken down to several basic themes of behaviors as in Figure 2.

Furthermore, we noted students’ concerns and feelings about specific classmates to juxtapose with (i) the reports from our collaborating teachers in order to finalize the group composition (described earlier), and (ii) the post-interview student data in order to identify potential changes in students’ relationships. Overall the pre-interview data confirmed the existence of conflict and bad emotions amongst teammates. In particular, we noted the reporting of physical and verbal conflict and bullying between specific teammates and associated numeric (Hate-Love) ratings.
being well below average on our 10-point Likert scale (mean = 5.75; SD = 3.06; N = 60). Pre-interview data also provided information regarding students’ access to technology. Although none of the students had seen a tabletop before, 18 (out of 20 students) had a computer at home and 16 of them had access to the Internet. Also, about half of the students had access to tablets or smartphones, either owned by them or their parents.

![Figure 2. Thematic network of the nature of conflict based on the pre-interview data](image)

**Post interview data**

We examined the post-interview data aiming to uncover potential changes in students’ relationships and persistent or non-persistent behaviors. Tracing students’ responses back to the pre-interview reports about specific teammates, it was evident that conflict between them was reduced and their relationships were improved. The dialogue excerpts below illustrate patterns of the positive outcomes of the intervention, with similar patterns traced for nearly all the participating students. It should be noted that during the post (also delayed) interview, students naturally spoke about their teammates, even though the interviewer asked more general questions about their daily school experiences and relationships with all their classmates. In a sense, students wanted to compare and be specific about the shifts in behaviors they had experienced.

**[Group 1, post-interview with AL]**

R: Tell me more about your current relationships with your classmates.
AL: Yesterday for example, KA and AI played hide-and-seek and I asked them to join. KA let me play with them!
R: Different than before?
AL: Yes, before he [KA] never let me play with them.
R: Are you telling me that you noticed a change in KA's behavior?
AL: Yes, he [KA] is nicer to me recently.
R: Do you think this has to do with you being in the same team for the recent tabletop activities?
AL: Yes I think so. Things with MO are also better because of this, I think…I think she [MO] is nicer after we played together on the tabletop.
R: Do you collaborate in class? outside of class?
AL: Both. During recess we play together and with some other friends too.
R: Any fights with her [MO]?
AL: Rarely… we are friends now…she [MO] changed…her behavior has changed, perhaps because we were in the same team…

**[Tracing back to the pre-interview with AL with reference to the same students].**

**[AL talks about KA]**

R: Tell me more about this issue with KA ….what is the problem exactly?
AL: I don’t know. He [KA] simply does not like me.
R: He [KA] does not like you….Perhaps there something you both like/have in common?
AL: Nothing.
R: There is nothing the two of you have in common?
AL: Absolutely nothing. I give him [KA] my pencils and I help him in class, but he never does the same. The other day I asked for his sharpener and he was screaming at me. And during the physical education class, the other day, he hit me… 2 or 3 times he threw me on the floor.
R: Was it on purpose?
AL: It is always on purpose.
R: How do you feel about this?
AL: I am sad…i want to be friends but he [KA] does not like me…I don’t know why.
.....

[AL talks about AI]
R: Give me an example of how you don’t get along with her [AI].
AL: For example, one day I sat next to her and she asked me to leave.
R: She [AI] prefers to sit with another friend?
AL: She just does not want ME!
R: But did she [AI] ever say so?
AL: But she never talks to me. The only thing she says is asking me to leave.
R: How do you feel about this?
AL: I want to play with them, but…
R: Do you think there is something AI and you have in common?
AL: Nothing.

[AL talks about MO]
R: Some days she [MO] likes you, some days not?
AL: Yes and I asked her why she does this to me, and she [MO] said I am a mean person.
R: Did you even do anything mean to her?
AL: No, I think it is something else…she does not like me.

Furthermore, we aimed to map students’ perceived value of the tabletop experience. A thematic network analysis (see Figure 3) showed that students’ reactions were overwhelmingly positive. The tabletop experience was perceived successful in bringing them closer together and in helping them realize that collaboration between them is possible. Also, in students’ talk about their experiences we could trace evidence of the affordances of tabletops to support collaboration and shared control, as illustrated below.

[Group 1, post-interview with KL]
R: So what you are telling me is that your group collaboration was always good except in one case when you and “AL” fought because she wanted to dominate?
KL: Yes, but everyone played in the end…Everyone could touch the tabletop because it was large and multitouch.
.....

[Group 3, post-interview with AN]
R: Did you work together as a group?
AN: Yes, we collaborated and if a teammate needed some help the rest of us would help. Last time for example [referring to IdeaSpace], one teammate started the story, another teammate continued, and then another one added more to it, until we had a very good story. It worked very well because we did it together.
R: So there was good collaboration? And everyone participated?
AN: Yes, everyone.
[...]
AN: I think the tabletop experience brought us closer together. Before this, we either did not talk or we talked badly to each-other. Now we play together.
R: Didn’t you play together before this experience?
AN: No, we never played before.
R: And what else do you do together now?
AN: We play football and volleyball during recess. We draw together during Art class and we do various other tasks together when the teacher assigns group work.
R: And do you fight at all?
AN: Not any more. [...] On the table, all tasks required collaboration; I think this helped us come closer together and learn to support each other in order to complete the tasks.

*Figure 3. Thematic network of the perceived value of the experience based on post-interview data*

Last, using the numeric data of the “1: Hate to 10: Love” response scale, a t-test was carried out to explore potential differences in students’ feelings about their teammates from pre to post testing. The results showed statistically significant increase in students’ “Love”, $t(59) = 5.7, p < .001$, from pre-testing (mean = 5.75, SD = 3.06) to post-testing (mean = 8.51, SD = 2.7). This difference constitutes a large effect size, Cohen’s $d = .96$ (see guidelines by Cohen, 1992), indicating that it is meaningful and may have practical importance (LeCroy & Krysik, 2007).

**Delayed interview data**

Last, we examined the delayed-interview data aiming to understand whether our intervention had long-lasting effect (3 months later). However, in this case the results were split and contradictory. With respect to relationships, students’ views can be summarised in three themes: (a) approximately one-third of the students reported good relationships with their teammates, (b) another one-third explained that they had good relationships with some teammates and not with others, and (c) another one-third disclosed information about conflict, similar to what we had recorded in the pre-interviews. With respect to recalling the experience and its value, results were again contradictory and various views were heard, regardless of the students reporting being in conflict or not at the time of the delayed interview. These views can be summarised in four themes: (a) some students believed the tabletop experience played a major role in their (at the time) good relationships, (b) others argued it helped temporarily but did not last for long, (c) others thought it did not help, even though they had a more positive view at the time of the post interviews, and (d) others were confused and unsure about the effects of the intervention. For example, in the case of AL (presented earlier), at the time of the delayed interview things were better with one teammate (MO) but not with the other two (AI and KA) and she felt the effect of the intervention did not last (see 1st excerpt below). Then, in the case of AN (also presented earlier) she enthusiastically elaborated on details of the tabletop experience and how it improved her relationship with her teammates (see 2nd excerpt below).

[delayed-interview with AL, Group 1]

R: How are your relationships with the other children these days [three months later]?
AL: With MO things are good. We sometimes we argue about things, but nothing major.
R: What kinds of things do you argue about?
AL: I don’t remember…it is small things (laughter).
R: With AI, KA? In the classroom? Outside the classroom?
AL: With KA we have problems.
R: Problems like in the “old days”? Big fights?
AL: Yes ...we totally don’t get along.
R: Can you give me an example?
AL: The other day, by mistake I gave the answer aloud to a question he [KA] was about to respond (the teacher asked him). During recess he hit me and made me cry... even though I had apologised he still wanted to hit me.
R: And with AI?
AL: Same thing. When I try to speak with her [AI], she screams at me and she kicks me. She does not like me.

[delayed-interview with AN, Group 3]

R: And do you remember the experience when you collaborated with your teammates…
AN: Yes I do. It was an educational table which helped us collaborate without actually noticing we were coming closer to each-other this way. It kind of promoted friendship between us, and it is lasting till today, and it was social, playful and fun.
R: But is it the tabletop really? Perhaps time just healed your relationships. Perhaps something else happened which helped…think about it a bit…
AN: I don’t know …but I do think it [the intervention] played a major role.
R: How? It was such a short experience…
AN: All games and activities on the tabletop required collaboration. One member did one thing, the others had to help him. For example, in the puzzle we should help each other find pieces quickly because time was running out and we wanted to win. And this made us work together and learn how to do this nicely (laughter).

Last, considering the numeric data of the “1: Hate to 10: Love” response scale, we found a statistically significant decrease in students’ “Love,” \( t(59) = 4.5, p < .001 \), from post-testing (mean = 8.51, SD = 2.7) to delayed-testing (mean = 6.23, SD = 3.26) with a medium to large effect (Cohen’s \( d = .76 \)). Yet, there was still a statistically significant increase in students’ “Love,” \( t(59) = 3.7, p < .001 \), from pre-testing (mean = 5.75, SD = 3.06) to delayed-testing (mean = 6.23, SD = 3.26), with a small effect size (Cohen’s \( d = .15 \)).

**Video data**

We examined the episodes of each group working on the same application across intervention sessions. As expected per design of the peacemaking activities and considering the literacy level of the participants, verbal communication was limited to short and simple conversations. Yet, together with the body language (including gestures, posture and movement), facial clues and tabletop interactions, it was possible to extract important information about the within-group collaboration over time. Overall, it was apparent that the technology facilitated students’ communication and collaboration as they took turns in trying inputs and completing the tasks together. In general, students engaged with the activity and cooperation did not appear to be forced or artificial. The investigators were able to identify patterns of behaviors across time, evident across all five groups. These theme-behaviors were consistent with the findings from students’ post-interviews (Figure 3), thus serving as a form of triangulation of findings:

- Conflict was evident at times, especially at the beginning of the intervention. Typically, teammates not involved in the particular conflict helped in order for the activity to proceed peacefully.
- From session to session collaboration got better in all groups. Typically after the first couple of sessions, students realized the importance of working together and appeared more willing to collaborate, take turns, consider everyone’s input and help each-other in order complete the task successfully.
- The technology allowed everyone to be involved and engaged in the activity. Dominant moves existed at times, but the multitouch capability allowed everyone to persist and continue to play/contribute. Typically after the first couple of sessions, students learnt to participate fairly and equally within their groups, demonstrating turn taking behavior.
- Students’ behavior was characterized by enthusiasm, engagement and playfulness, suggesting this was a positive experience for all. None wanted this to end, and at the end of each session, everyone asked to participate in the next session.
Furthermore, the video analysis allowed us to take a closer look at how each activity facilitated peacemaking differently. For example, the use of iMath initially promoted competition within the group and encouraged the strongest players in math to dominate the activity by choosing the right answer before anyone else could think. In the early sessions of the intervention, this practice enforced conflict in the group, particularly between individuals being already in conflict. Soon however, a turn taking behavior was adopted by all groups, when students realized that (i) the tabletop was picking the wrong answer when everyone touched concurrently and aggressively therefore points were lost, and (ii) time lost in arguments and fighting was counting against the group winning the game. To provide a second example, the use of IdeaSpace encouraged students, especially the less talkative ones, to share their feelings and emotions. In particular, during the storytelling activity, some students related themselves to actors of their story and expressed thoughts which, rather unconsciously, communicated a message to their teammates. Also, these expressive moments revealed aspects of the students’ family situation (for which the investigators were already informed), suggesting further indirect benefits of the activity. For example, in a student’s own words: “Who wants their mom and dad to be fighting? Who wants to be screamed at all the time, at home and at school? This boy in our story is probably very sad. We need to add some friends for him in the story and a teacher who tolerates and excuses his behavior.” [XO, Group 5]

A detailed analysis of how our six tabletop activities supported peacemaking in different ways is beyond the scope of this manuscript and is considered elsewhere.

Discussion

The potential of multitouch interactive technologies for peacemaking is widely unexplored to date. This is the first investigation of collocated technology-enhanced peacemaking with young student-participants in a real conflict-stressed school environment. Our findings suggest that collocated collaboration around tabletops can mediate peace within conflict-laden groups. We have evidence that the technology becomes a means for communication and collaboration giving the chance for students in conflict to share a common space, shifting attitudes and improving their relationships. Below we discuss these findings in relation to prior research, while pointing to future possibilities.

First, our findings suggest that tabletops can promote collocated collaboration among students in conflict. It is not new that tabletops afford collocated collaboration and teamwork within groups (e.g., Dillenbourg & Evans, 2011; Morris et al., 2006), but our work takes a step further to demonstrate collaboration among students in conflict. In previous peace-related work, the NNR-Table (Stock et al., 2008; Zancanaro et al., 2012) promoted dialog and understanding of culture between participants from two opposing sides. Yet, the adolescent participants, did not exhibit conflict between them per-se. Instead their volunteered participation in the study might be suggesting the opposite, i.e., their interest in reconciliation and peace. Similarly, in the case of PeaceMaker (Burak et al., 2005) the participants did not know each other personally prior to the study, i.e., there was no personal conflict between them. In the present investigation, we worked with young (elementary-school) students in conflict, forming conflict-laden groups nominated by their teachers and confirmed by the students themselves. Our study showed evidence that collocated collaboration around tabletops helps students in conflict share a common space and shape better relationships among them.

Second, we demonstrated that tabletops are well suited for collaborative activities in case of low literacy levels and communication barriers. Previously, Rick et al. (2011) argued that gesture-based communication around tabletops supplements, and even replaces, verbal communication. The present investigation confirms this argument. We observed fluent collaboration over limited verbal communication (supported by non-verbal communication) within all the participating groups. We acknowledge however that in IdeaSpace, teacher facilitation was needed as narrating a story was challenging for the students. This is consistent with Buckner and Kim’s (2011) reports about low literacy levels being a barrier to students’ ability to read, write, and share their stories during peacemaking interventions using mobile devices. Yet, all other activities in our intervention (e.g., those with more game-like characteristics) were completed successfully without facilitation. Therefore, we can argue that tabletops, enriched with game-like activities, afford gesture-based communication and collaboration over language barriers and therefore, become well suited for collocated peacemaking interventions when language barriers to communication exist.
Moreover, we suggest that tabletops are well suited for peacemaking school interventions because they help create a playful learning environment. Playfulness around tabletops has already been reported in the literature (e.g., Pieter, Friedman, & Hollan. 2012; Ioannou, Zaphiris, Loizides, & Vasiliiou, 2013; Jamil et al., 2011). For example, Ioannou et al. (2013b) found no evidence of tension, anxiety or strong disagreement when college-students discussed sensitive issues around a tabletop; the authors discussed how the playfulness of the tabletop interaction, as documented in utterances of social talk and laughter, had possibly made the activity fun, rather than emotional. In the present investigation, through self-reports and video observations, we gathered evidence that the intervention was fun, playful and enjoyable for all the participants despite the existence of conflict between them. Thus, not only do we confirm previous research findings about tabletops promoting playful interactions, but we further argue that this affordance makes them well suited for peacemaking interventions. In a broader sense, we can argue for the potential of using tabletops for gamifying the peace education curricula activities.

Furthermore, our findings extend the notion that tabletops are “peace technology” in that they allow power to be shared and distributed over the display enabling participation by all collaborators (e.g., Fleck et al., 2009; Ioannou, Zaphiris, Loizides, & Vasiliiou, 2013; Marshall et al., 2008) and even “forcing” the participants to consider the interests of the “other” (Ioannou et al., 2013b). As described earlier, through video analysis, we noted the adoption of turn taking behavior in all groups, which suggests virtually equal participation between the collaborators. Moreover, students’ self-reports informed how the technology allowed everyone’s input to be considered, preventing one from dominating the activity, which also suggests a more democratic participation to the activity. Veletsianos and Eliadou (2009) argued that for peace-seeking initiatives to be effective, it is necessary that they promote an understanding of the “other.” We acknowledge that because tabletops enabled democratic participation, it does not mean an understanding of the “other” took place. For this goal to be achieved, we argue that a stronger pedagogical aspect aimed explicitly at social perspective taking is needed (e.g., Gehlbach et al., 2008; Goldsworthy et al., 2007).

Last but not least, we could argue that infusion of tabletops in the general curriculum might be able to help promote peace-related outcomes. To elaborate, tabletop use in this study expanded on the math, English, and general education curricula to turn attention away from peacemaking and life education as formal curricula activities. This design decision seems to reveal a possibility – that activities and applications explicitly designed for peacemaking (e.g., PeaceMaker, Burak et al., 2005) or dedicated peacemaking interventions might be less needed. Instead, technology infusion in the general curriculum, can feed into peacemaking by targeting “antecedents to peace” (Veletsianos & Eliadou, 2009), such as supporting and mediating collaboration, interaction and communication. For example, using tabletops in everyday collocated collaborative activities (math, science, English or other) might allow students in conflict to work together and experience shifts in attitudes such as those reported here. In this sense, tabletops can “expand” the peace education curricula, which is often under-emphasized, helping to promote peace-related outcomes through their infusion in the general curriculum. We acknowledge these ideas are preliminary and warrant further investigation.

With respect to the limitations, we acknowledge that much enthusiasm was evident in students’ feedback, stemming from the use of tabletops. Although this finding relates to students’ endorsement of the tabletop experience (e.g., Buisine et al., 2012; Ioannou, Zenios, & Stylianou, 2014; Ioannou, Christofi, & Vasiliiou, 2013a), it might also be suggesting a novelty effect, which can be confused with the true impact of the intervention. A novelty effect becomes particularly relevant when we consider the variability in the results of the delayed-interview data, with a few students returning back to the baseline with regard to their relationships and feelings about their teammates. In their review, Veletsianos and Eliadou (2009) found that there is generally limited empirical support for the sustainability of the claimed positive outcomes of various technology-infused peace initiatives. Indeed, collecting delayed-interview data in this work was an attempt to address this concern, nonetheless the mixed and conflicting results leave us with no clear picture of the lasting effects of the intervention. We therefore argue that longitudinal studies using ethnographic approaches are needed, in order to examine issues of “depth” related to technology-enhanced peacemaking and sustainability of any positive outcomes. Furthermore, like it is also true with many new and emerging technologies, theorizing the use of tabletops for peacemaking, but also for collaborative work/learning in a broader sense, is in very early stages. Future work should focus on presenting a theoretical account of how tabletops might promote peace-related outcomes by eliciting and supporting various types of interactions.
Conclusions

Overall, this study presented evidence that tabletops are “peace technology” that is well suited for use in peacemaking interventions aiming to promote communication and collaboration among students in conflict. This is the first empirical work of collocated technology-enhanced peacemaking in school contexts with great promise for future work. Our findings should be informative for researchers and educators dealing with phenomena of school conflict, as well as researchers and practitioners in the broader area of using technology for peace and social change.

We summarize a few tentative implications of this work for these stakeholders:

- Tabletops can promote collocated collaboration among students in conflict.
- Tabletops enable democratic participation and consideration of everyone’s input.
- Tabletops enable communication over language barriers.
- Tabletop can help create a playful learning environment.
- Technology interventions that involve collocated users, present the potential to help change perceptions, improve relationships and break down social barriers.

Closing, tabletops are a new generation of educational technologies that offer new possibilities for engaging students in communication and collaboration. New forms of collaborative practice can be supported by this technology, especially as it gets more advanced and equipped with relevant software applications. What is also important, this kind of technology is attractive for young people who stay engaged and enjoy the experience. Tailoring this technology to meet the need for peaceful attitudes and peacemaking in conflict-stressed environments is a great area for research and development while serving an important purpose. We believe that tabletops provide a revolutionary approach to collocated technology-enhanced peacemaking in school contexts and beyond. Its integration in education, might give us the opportunity to deal with social issues in ways we have not had the chance to do before.

References


Exploring Engaging Gamification Mechanics in Massive Online Open Courses

Jen-Wei Chang1* and Hung-Yu Wei1, 2

1Graduate Institute of Communications Engineering, National Taiwan University, Taiwan // 2Department of Electrical Engineering, National Taiwan University, Taiwan // jenweichang@ntu.edu.tw // hywei@cc.ee.ntu.edu.tw

*Corresponding author

(Submitted October 2, 2014; Revised June 30, 2015; Accepted September 13, 2015)

ABSTRACT

Massive open online courses (MOOCs) have developed rapidly and become tremendously popular because of their plentiful gamification designs, such as reputation points, rewards, and goal setting. Although previous studies have mentioned a broad range of gamification designs that might influence MOOC learner engagement, most gamified MOOCs fail to meet learning objectives because of a lack of research regarding suitable game design, as well as poor rationale for or design of gamification mechanics. This study aims to explore and identify engaging gamification mechanics for MOOC learners. We conducted a focus group interview with 25 MOOC frequent users to identify 40 gamification mechanics. This study then determined the relative engagingness of these gamification mechanics by administering an online survey to 5,020 MOOC learners. The results indicated that the 10 most engaging gamification mechanics accounted for more than 50% of the engagingness. The mechanics of the Where’s Wally game is extremely engaging for MOOC learners; however, they it is not been demonstrated in previous relevant studies. Finally, we discuss the top five engaging gamification mechanics and their implications.

Keywords
Improving classroom teaching, Interactive learning environments, Gamification

Introduction

Massive online open courses (MOOCs) are a current trend for creating online courses for equipping learning institutions to obtain a free and high quality teaching initiative with relevant visibility on the Internet (Johnson, Becker, Cummins, Freeman, Ifenthaler, & Vardaxis, 2013; Pellas, 2014). MOOCs refer to web platforms that allow millions of learners to access various instructional materials and resources without the constraints of time and place, and additional learning opportunities to supplement traditional classroom instruction, such as Coursera, Udacity, and edX (Lin, 2010; Stoel & Lee, 2003). MOOCs are interactive, online learning tools that support the learning of specific concepts by enhancing, amplifying, and guiding the cognitive processes of learners (Altbach, 2014). MOOCs use the increasing popularity of social networking services (SNSs) such as instant messengers (IMs), Facebook, and Twitter, to facilitate increased social interaction and engage millions of teachers, learners, and parents (Lin, 2010). The learner–learner and learner–instructor interaction created by MOOCs is central to knowledge acquisition and the development of learner cognitive skills, and that interaction is intrinsic to effective instructional practice (Lee & Hammer, 2011; Tobarra, Robles-Gómez, Ros, Hernández, & Caminero, 2014). MOOCs are an alternative to traditional models of face-to-face education, and have even been viewed as a threat to traditional educational institutions and professionals (Millard, Borthwick, Howard, McSweeney, & Hargood, 2013). Thus, the development of MOOCs has received considerable attention from both educators and learning-technology developers.

MOOCs have been an increased focus related to learner participation of MOOCs, given rising tuition costs and concerns regarding learner success and retention rates (Pappano, 2012). Although MOOCs are rapidly developing and gaining enormous popularity, most of them fail to help learners to remain focused on learning content and lead to relatively poor learning efficiency and effectiveness. This phenomenon occurs because most MOOC designs do not provide learners with an engaging experience. Certain researchers have mentioned that MOOCs must enhance learner digital engagement, which refers to the learning and everyday engagement of learners with available technologies in their learning ecologies, including both daily life and school contexts (Gurung & Rutledge, 2014). Therefore, improving learner digital engagement is critical to the development of MOOCs.

Certain studies have proposed gamification as a potential solution to alleviate this problem (Grünewald, Meinel, Totschnig, & Willems, 2013; Skiba, 2013; Dicheva, Dichev, Agre, & Angelova, 2015). Gamification incorporates game mechanics into nongame settings to increase user engagement and enjoyment of a product or service, and to
encourage users to perform certain behaviours (Hsu, Chang, & Lee, 2013). Gamification essentially functions as entertainment that causes learners to enjoy actively participating and engaging with others, such as through reputation points, rewards, and goal setting. Kapp (2012) indicated that gamification is crucial to the development of learning technology because numerous elements of gamification are based on educational psychology and techniques that instructors have been using for years. Simões, Redondo, and Vilas (2013) developed a learning platform for K-6 learners, and suggested that education is an area with high potential for applying gamification because it substantially promotes learner motivation and engagement with the learning platform. Sung and Hwang (2013) proposed a gamification mechanism for course websites to improve the learning performance of learners in their learning attitudes, learning motivation, self-efficacy, and learning achievements. Because of the importance of gamification to learner engagement, certain popular MOOCs such as Coursera, Udacity, and edX effectively attract and maintain learners through various gamification designs such as rewards and badges. These gamification design factors form social engagement loops by providing fun and flow experience as learners interact with websites, which result in more daily visitors and a higher average time spent on sites (Zichermann & Cunningham, 2011). Consequently, gamification plays a critical role in the success of MOOCs.

Previous studies have mentioned numerous gamification design factors for MOOCs, which can be classified into three types of interactivity, as proposed by Moore (1989) (see Figure 1): (a) learner–content interaction, (b) learner–learner interaction, and (c) learner–instructor interaction. Numerous studies have examined the role of learner–content interaction (such as time pressure and status) in facilitating learner engagement through interacting with the subject matter under study to construct meaning, relate it to personal knowledge, and apply it to problem solving (Reeves & Read, 2009; Deterding, Sicart, Nacke, O’Hara, & Dixon, 2011). Some scholars have argued that gamification designs should rely heavily on the mutual support and socializing of learner–learner interaction (such as peer-tutoring and group identification) because the peer group relationship can enhance regular participation (Choi, & Kim, 2004; Williams, Ducheneaut, Xiong, Zhang, Yee, & Nickell, 2006; Chen, Sun, & Hsieh, 2008; Jang & Ryu, 2011; Hou, 2012; Lee & Chang, 2013). Other studies have examined the learner–instructor interaction process that stimulates, enhances, and maintains learner engagement with a subject (such as rewards and goal setting; Ryan & Deci, 1996; Ducheneaut & Moore, 2004; Hsu, Wen, & Wu, 2009).

Although the aforementioned studies indicated a broad range of factors that might influence gamification, they depicted neither the actual design mechanics nor the relative engagingness among them. Gamified MOOCs fail to meet learning objectives because of a lack of research regarding suitable game design, as well as poor rationale for or design of gamification mechanics. The selection of crucial gamification mechanics is a multiple-criterion decision-making (MCDM) problem. Previous studies (Kahraman, Cebeci, & Ulukan, 2003; Büyüközkan, 2004; Kim & Nevo, 2008) have regarded the analytic hierarchy process (AHP) as an appropriate method for solving these MCDM problems. Decision makers have to decompose the goal of the decision process into its constituent parts, progressing, from the general to the specific perspective. The structure of AHP has to include a goal, criteria and alternative
levels, ordered into a hierarchy. Each item (criterion, sub-criterion or alternative) would be divided into an appropriate hierarchy of detail. Specifically, decision makers judge the importance of each criterion in pair-wise comparisons, structured in matrices. The scoring of AHP is on a relative basis, comparing the importance of one decision alternative to another. This study therefore adopted a hybrid methodology combining fuzzy logic techniques and the analytic hierarchy process (FAHP) approach. This hybrid methodology provides a systematic tool for analysing learners’ sense of relative engagingness gamification of gamification mechanics and assists decision makers in decomposing the multi-criteria problem into a hierarchical model.

Identifying the engaging gamification mechanics can establish critical milestones in how to create highly effective MOOCs. This study aims to explore the gamification mechanics of MOOCs and determine the relative engagingness of these gamification mechanics. The results can assist MOOC website designers in designing highly engaging MOOCs. The results also identify the engaging gamification mechanics for instructors to enhance learner’s engagement.

Gamification in interactivity

Interactivity has been defined differently, each definition reflecting the perspectives of the group using it. Weller (1988) describes interactivity as an event or a process that occurs when a learner actively adapts to information being presented by a form of technology that, in turn, adapts to the learner. Merrill, Li, and Jones (1990) argue that interactivity in learning involves real-time dynamics and mutual give-and-take between an instructional system and a learner—especially in relation to exchanges of relevant information. Apparently, these definitions address interactivity’s accounting for the relationships between a learner and the instructional content presented by either an instructor or an instructional system.

All gamification factors discussed in this study includes three main components: learner–content interaction, learner–instructor interaction, and learner–learner interaction. Learner–content interaction implies learners interacting with the subject matter under study to construct meaning, relate it to personal knowledge, and apply it to problem solving (Reeves & Read, 2009; Deterding et al., 2011). Learner–instructor interaction refers to stimulating, enhancing, and maintaining learner motivation (Ryan & Deci, 1996; Ducheneaut & Moore, 2004; Hsu et al., 2009). Learner–learner interaction refers to interaction among individual learners or among learners working in small groups (Choi & Kim, 2004; Williams, Ducheneaut, Xiong, Yee, & Nickell, 2006; Chen, Sun, & Hsieh, 2008; Jang & Ryu, 2011; Hou, 2012; Lee & Chang, 2013).

After reviewing gamification factors from literature, we conducted an in-depth interview with five MOOC experts who had more than 3 years of MOOC developing experience. Therefore, through this process, certain factors were excluded because they lacked corresponding applications for MOOCs.

Learner–content interaction

Self-expression

Self-expression refers to peoples’ desire to express their autonomy and originality, which shapes their unique personalities (Hsu et al., 2009; Antin & Churchill, 2011). Learner self-expression involves a feeling of social toleration, life satisfaction, public expression, and an aspiration to liberty. Gee (2003) conducted a study on digital-game-based learning and considered that assisting learners to build their self-identity in a virtual world can facilitate learner engagement.

Pattern recognition

Pattern recognition refers to the dynamics of learner-content interaction most associated with unpacking website complexity (Zichermann & Cunningham, 2011). When learners seek to understand the composition of learning content and explore hidden meanings or how complex items interact, they are seeking pattern recognition. When
patterns are detected, learners organize the learning content around those patterns, and typically feel intrinsically rewarded simply for having discovered them.

**Time pressure**

Numerous game designs use time as a motivator for player activity and action (Reeves & Read, 2009; Antin & Churchill, 2011; Hsu et al., 2013). Time pressure means MOOCs give learners a time limit to perform certain learning behaviours to encourage them to interact heavily or to complete necessary tasks. For learners, creating time pressure can arouse more emotional feedback and encourage greater participation because the time pressure is connected to their goals. For example, certain learning game applications set a 5-s time limit to find targets, which encourages users to interact heavily with the application during this period. When they fail, a new game automatically begins 5-s later.

**Status**

When learners join a social group, status refers to learners’ need for recognition, fame, prestige, attention, and other learners’ respect (Antin & Churchill, 2011). Status serves as learners’ desire for recognition and encourages learners to achieve goals enthusiastically. For MOOCs, status also represents each learner’s contribution to course resources and participation in learning activities. Quantified evaluation is frequently used for representing design mechanics regarding learner status.

**Learner–instructor interaction**

**Goal setting**

Goal setting is related to the most motivating goals, which are those that are just out of comfortable reach (Lin & Chang, 2005). In the learning environment, learners are motivated to pursue a specified goal because goal seeking itself is often the primary reward (Antin & Churchill, 2011). Learner goals can comprise personal level goals or group level goals.

**Instruction**

When new learners (also called newbies) enter a system, certain instructions are required to teach them social norms (Montola, Nummenmaa, Lucero, Boberg, & Korhonen, 2009; Antin & Churchill, 2011). Instruction functions as the social shaping of learning activities and assists learners in mastering an entire system efficiently. Instruction is often used for debriefing and offering feedback so that learners can understand what occurs in a learning system and how these events support the instruction objectives (Kapp, 2012). In the context of MOOCs, instruction assists learners in learning communication and teamwork skills as they collaborate with others.

**Rewards**

Rewards refer to the gamification factors that satisfy learners’ shared need and motivate them to engage in learning activities (Ryan & Deci, 1996). For example, learners are motivated to perform additional problem-solving behaviours to receive additional rewards from websites. The reward mechanism operates by awarding points or equivalents (e.g., frequent-flyer miles) and effectively forms a reward-behavior cycle (Hsu et al., 2009). Learners who invest more time in the encouraged behaviours receive more from the learning system. Rewards can be classified into intrinsic and extrinsic rewards according to their motivation. Intrinsic rewards allow learners to engage in learning for greater self-fulfillment, and extrinsic rewards allow learners to learn for earning something (Lee & Hammer, 2011).
Learner–learner interaction

Reputation points

Reputation points are a mechanism that encourages learner behaviours based on the estimation of recognition held by others inside and outside of an organization (Tulathimutte, 2006; Wolf, 2007). The concept of reputation points has been commonly adopted on online shopping websites such as eBay and Amazon.com to increase system reliability, reduce risks between users, and assist users in deciding whether to interact with and trust a user based on the experiences of other users with that user. Several online games, such as World of Warcraft (WOW) and Ultima Online (UO), use reputation points to recognize users who have fought with other players of comparable experience levels to obtain special titles and items. Learners’ desire for reputation points can be considered a motivation for engagement because they play harder to increase their reputation in the game.

Peer tutoring

A peer tutor is anyone who is of a similar status to the person being tutored (Höysniemi, Hämäläinen, & Turkki, 2003; Huang, Yeh, Li, & Chang, 2010). In an undergraduate institution, this is typically other undergraduates, distinct from the graduate students who might be teaching writing classes; in a K-12 school, this is typically a student from the same grade or higher.

Competition

Competition refers to a learner’s desire to compete with others, including reaching a higher score and winning over others (Yee, 2006). When a learner competes with others, the learner with the highest score wins a prize or other benefit. Thus, learners enjoy the well-being and continue competing with others.

Altruism

Altruism is a learner’s desire to form and maintain relationships with others through certain behaviours, such as gift-giving or asking for help (Antin & Churchill, 2011). Trivers (1971) suggested that altruism is a learner’s desire to conduct reciprocal behaviours with others based on trust. Altruists indirectly contribute to their fitness through others who reciprocate. In the MOOC environment, support for gift-giving and charity is the most popular altruistic. Altruism is also considered a strategy to attract new learners (Antin & Churchill, 2011). For instance, learners can receive a gift from someone that draws them into the MOOCs, and are subsequently motivated to send gifts to other learners for reciprocity purpose, eventually creating a great acquisition loop.

Group identification

Group identification represents learners’ affective and cognitive loyalty to a learner group (Lee & Chang, 2013; Bergami, & Bagozzi, 2000; Jo, Moon, Garrity, & Sanders, 2007; Pisan, 2007). Learners with higher group identifications are often willing to remain in a group permanently and to strive toward goals, obey the guild manager’s commands, and devote themselves to group affairs (Seay, Jerome, Lee, & Kraut, 2004).

Peer appraisal

Peer appraisal has been historically used for logistical, pedagogical, metacognitive, and affective benefits (Sadler & Good, 2006; Conejo, Barros, Guzmán, & García-Viñas, 2013), and offers a promising solution that can scale the grading of complex assignments in courses with tens or even hundreds of thousands of students. When using MOOCs, instructors cannot review essays or other open-ended work from thousands of students as they do in smaller class settings. To remove this limitation, MOOC providers are looking to peer-based assessments, in which students learn to review the work of their cohorts.
Method and result

Identifying gamification mechanics for MOOCs

Materials

The most highly subscribed MOOCs, Coursera, Udacity, and edX, were surveyed in this study. These three MOOCs represent online learning platforms and have myriads of users worldwide.

Subjects and procedures

Twenty-five frequent users including fifteen distance-learning course teachers, four distance-learning students, and six MOOC developers were invited to participate in the interview. All of them have more than two years’ experience in using MOOCs. Participants were asked to identify and discuss gamification mechanics of MOOCs based on the gamification factors. The interview questions and record are in the Appendix 1.

Result

To improve gamification mechanics, three human–computer interaction experts with more than 7 years of experience were invited to confirm all of the garnered mechanics. Finally, 40 gamification mechanics were developed following the in-depth interviews, as shown in Appendix 2. To depict the relationship between gamification elements and demonstrate and different levels, a concept map was proposed in Figure 2.

Figure 2. The concept map of all gamification mechanics in MOOCs
Determining the relative engagingness of gamification mechanics

Questionnaire design and data collection

Shown in Appendix 3, AHP questionnaire was developed to gather MOOC learner assessments of the relative engagingness of the gamification mechanics in a pairwise comparison-data input format. Each item was assessed using a 9-point ratio scale, as suggested by Saaty (1990). Each item was scored using a scale comprising equally engaging, moderately engaging, strongly engaging, very strongly engaging, and extremely engaging. This study conducted an online survey to gather data. The survey was advertised on four MOOCs and e-learning online forums in Taiwan to recruit volunteers to participate in this study. After excluding volunteers with incomplete data, 5020 users’ data were collected in this study. Table 1 summarizes demographic data of all subjects. Following primary data analysis, we deleted incomplete questionnaires and outlier data, leaving 4,891 valid samples (97.43%) for use in this study.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Participants (N = 296)</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2437</td>
<td>48.55%</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2583</td>
<td>51.45%</td>
<td></td>
</tr>
<tr>
<td>Frequently-used MOOCs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coursera (<a href="https://www.coursera.org/">https://www.coursera.org/</a>)</td>
<td>2514</td>
<td>50.08%</td>
<td></td>
</tr>
<tr>
<td>Proera (<a href="http://www.proera.com.tw/">www.proera.com.tw/</a>)</td>
<td>2216</td>
<td>44.14%</td>
<td></td>
</tr>
<tr>
<td>Taiwan open courseware (<a href="http://www.tocwc.org.tw/">www.tocwc.org.tw/</a>)</td>
<td>2057</td>
<td>40.98%</td>
<td></td>
</tr>
<tr>
<td>ewant (<a href="http://www.ewant.org">http://www.ewant.org</a>)</td>
<td>1782</td>
<td>35.50%</td>
<td></td>
</tr>
<tr>
<td>Open edX (<a href="https://courses.openedx.tw">https://courses.openedx.tw</a>)</td>
<td>1521</td>
<td>30.30%</td>
<td></td>
</tr>
<tr>
<td>Share Course (<a href="http://www.sharecourse.net/">http://www.sharecourse.net/</a>)</td>
<td>1305</td>
<td>26.00%</td>
<td></td>
</tr>
<tr>
<td>Taiwanlife (Taiwanlife.org)</td>
<td>1227</td>
<td>24.44%</td>
<td></td>
</tr>
<tr>
<td>NTU MOOC (<a href="http://www.ntumooc.org/">http://www.ntumooc.org/</a>)</td>
<td>963</td>
<td>19.18%</td>
<td></td>
</tr>
<tr>
<td>Udacity (<a href="https://www.udacity.com">https://www.udacity.com</a>)</td>
<td>802</td>
<td>15.98%</td>
<td></td>
</tr>
<tr>
<td>edX (<a href="https://wwwedx.org">https://wwwedx.org</a>)</td>
<td>775</td>
<td>15.44%</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>750</td>
<td>14.94%</td>
<td></td>
</tr>
<tr>
<td>MOOCs use experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-12 months</td>
<td>655</td>
<td>13.05%</td>
<td></td>
</tr>
<tr>
<td>12-36 months</td>
<td>2983</td>
<td>59.42%</td>
<td></td>
</tr>
<tr>
<td>&gt;36 months</td>
<td>1382</td>
<td>27.53%</td>
<td></td>
</tr>
<tr>
<td>Average age</td>
<td>23.02 years old (Std. = 2.11).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data analysis

Our proposed fuzzy-AHP approach included seven steps. We first used triangular fuzzy numbers to construct the fuzzy comparison matrix, as shown in Table 2. Second, we integrated the collected user assessments of each gamification mechanic, design factor, and design component by using the fuzzy average method proposed by Buckley (1985). Third, we computed the fuzzy weight of each gamification mechanic by using the approximation method introduced by Buckley (1985). Fourth, the center of gravity method, a defuzzifying method proposed by Tzeng and Teng (1993), was performed to defuzzify the weight of each gamification mechanic. Fifth, we normalized the weights of all gamification mechanics. Sixth, we aggregated each level of the proposed gamification framework and calculated the relative engagingness value of the fuzzy weight for each mechanic at factor levels. Finally, we computed the consistency index (CI) and consistency ratio (CR) for each fuzzy comparison matrix. The detailed process of data collection and the proposed fuzzy-AHP model are in Appendix 4.

<table>
<thead>
<tr>
<th>Table 2. Membership function and definitions of fuzzy numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuzzy number</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>
The weight of all gamification mechanics is shown in Appendix 5. Among the 40 engaging gamification mechanics, the top 10 most engaging mechanics, listed in Table 3, can account for more than 50% (51.68%) of engagingness. In our study, the CR was $0.072 \leq 0.1$, then the output of the pair-wise comparison can be proven sufficiently consistent.

**Table 3. Top10 Most engaging gamification mechanics in MOOCs**

<table>
<thead>
<tr>
<th>Gamification mechanics</th>
<th>Contribution (%)</th>
<th>Accumulative contribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM1: Virtual goods</td>
<td>9.52</td>
<td>9.52</td>
</tr>
<tr>
<td>GM23: Redeemable points</td>
<td>8.45</td>
<td>17.97</td>
</tr>
<tr>
<td>GM31: Team leaderboards</td>
<td>7.34</td>
<td>25.30</td>
</tr>
<tr>
<td>GM5: Where’s Wally game</td>
<td>4.76</td>
<td>30.06</td>
</tr>
<tr>
<td>GM13: Trophies and badges</td>
<td>4.61</td>
<td>34.67</td>
</tr>
<tr>
<td>GM38: Peer grading</td>
<td>4.15</td>
<td>38.82</td>
</tr>
<tr>
<td>GM40: Skill points</td>
<td>3.93</td>
<td>42.74</td>
</tr>
<tr>
<td>GM4: Memory-game interactions</td>
<td>3.31</td>
<td>46.05</td>
</tr>
<tr>
<td>GM8: Check points</td>
<td>2.89</td>
<td>48.94</td>
</tr>
</tbody>
</table>

**Discussion and implications**

In this section, we discuss the top five engaging mechanics following the aforementioned analysis. The practical implication for both instructor and MOOC website designer is also discussed (see Table 4).

**Table 4. The implication for Instructor and MOOC website designer**

<table>
<thead>
<tr>
<th>Top five gamification mechanics</th>
<th>Instructor</th>
<th>MOOC website designer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual goods</td>
<td>• Quantify learner’s contribution to earn virtual goods</td>
<td>• Virtual goods exchange interface</td>
</tr>
<tr>
<td></td>
<td>• Special challenge for special virtual goods</td>
<td>• Display virtual goods on personal page</td>
</tr>
<tr>
<td>Redeemable points</td>
<td>• Clear redeemable point’s rules.</td>
<td>• Dialogues should not contain information that is irrelevant or unnecessary.</td>
</tr>
<tr>
<td></td>
<td>• Integrate the redeemable points into the course content.</td>
<td>• Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.</td>
</tr>
<tr>
<td></td>
<td>• Allow learners use the points accumulated from MOOCs to redeem real world rewards e.g., course material, items, toys, and game software to increase course participation.</td>
<td>• Redeemable points status.</td>
</tr>
<tr>
<td></td>
<td>• Make simple and visible comparisons between learner’s teams.</td>
<td>• Reminder learners how far they can proceed to next level of rewards.</td>
</tr>
<tr>
<td>Team Leaderboards</td>
<td>• Quantify the team participation</td>
<td>• Local view: allow learners immediately see how they rank among their friends and classmates.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Global view: allows learners to see how they rank among all learners within the system as a whole.</td>
</tr>
</tbody>
</table>
Where’s Wally game  
- Link the problem solving process with Where’s Wally.
- Develop learner’s pattern recognition, critical thinking, and sense-making skills.
- Give obvious cues as learners are stuck.
- Design various faintly discernible cues.
- Display different cues according to learner’s skill level.

Trophies and badges  
- Immediately reward learner’s achievement using trophies and badges.
- Develop milestone badges to enhance learner’s motivation by collecting the badges.
- Trophies and badges ladder
- Display the trophies and badges information inside the learner’s personal page and information portion of the comment page.

Among the five most engaging gamification mechanics of MOOCs, virtual goods were the most engaging gamification mechanic. Such virtual gifts can be linked to learner’s achievement motivations. The result means learners engage more with the MOOC in order to earn greater achievement. This finding is consistent with Denny’s (2013) work. He reported on a large-scale experiment measuring the impact of virtual achievement in e-learning applications and found the virtual achievement has a positive motivational effect on learner’s engagement. His result also showed learners prefer earning and owning virtual goods. The practical implication for instructor is to quantify learner’s contribution and give learners virtual goods if they achieve a certain level.

Instructors can also provide special challenge for them to earn special virtual goods. Virtual goods work as positive reinforcement for learner’s good performance or regular participation. The practical implication for MOOC website designer is to provide a virtual goods exchange interface and display the earned virtual goods in the learner’s personal page.

The second most engaging gamification mechanic was redeemable points, which indicate the redeemable points engage learners by supporting their personal achievement motivation. This finding is consistent with Grünewald, Meinel, Totschnig, and Willems (2013)’s work. Grünewald and his colleague (2013) collected 2726 active MOOC participants’ data to investigate multiple learning styles and found redeemable points engage learners by strengthening the social incentives. From instructor’s perspective, the implication is to integrate the redeemable points into the course content. Instructors can allow learners use the points accumulated from MOOCs to redeem real world’ rewards e.g., course material, items, toys, and game software to enhance the course participation. From MOOC website designer perspective, they should design simple and clear redeem user interface to assist learners redeem gifts. Dialogues should not contain information that is irrelevant or unnecessary. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility. The redeem interface also have to display the redeemable points status and reminder learners how far they can proceed to next level of rewards.

The third most engaging gamification mechanic was team leaderboards, which means that learner-learner interaction such as the comparison and competition of teams receives more attention for MOOC learners compared with individual leaderboards. The visible competition among teams encourages team members’s highly engagement. This result is consistent with Chiu, Hsu, and Wang’s (2006) work. They examined knowledge sharing in a virtual learning community and observed that visible competition among teams substantially increase learners’ reliance on team effort rather than on their own effort, and thus engage learners to cooperate with team members. From instructor’s perspective, the implication is the instructor should quantify learner’s team participation and then encourage simple and visible comparisons between learner’s teams. The practical implication for MOOC website designer should provide two types of team leaderboards: local and global view. A local view allow learners immediately see how they rank among their friends and classmates. A global view allows the learners to see how they rank among all learners within the system as a whole, it will assist learners to check the actual learning performance and establish their learning goals.

The Where’s Wally game was the fourth most engaging gamification mechanic. This result means learners seek to understand the world around them, and attempt to discover the hidden meaning or complex items interact in MOOCs. The problem-solving process of the Where’s Wally game also provides strong motivation for learner engagement, a finding that has not been revealed in the gamification study. In previous learning studies, Where’s Wally is considered as a type of wordless figure and which is highly correlated to visual literacy for learners.
The wordless picture are considered as an effective learning tool which encourages readers learning behaviour including sense-making, problem solving, critical thinking, etc. (Avgerinou & Ericson, 1997). Jalongo et al. (2002) also considered the wordless pictures connect learner’s visual literacy skills, culture literacy (learning the characteristics and expectations of social groups) and literacy with print. Crawford and Hade (2000) investigated children’s reading of wordless picture books, found the children make sense of wordless picture books by using sense-making processes. They also found the wordless picture books let children construct meaning with prior knowledge and experiences, attention to intertextual cues, multiple perspective-taking, reliance upon story language and rituals, and the implementation of active, playful behaviours as part of the reading process. Therefore, we consider Where’s Wally game engage learners by connecting learner’s visual literacy, which contains critical thinking, learning, construct meaning, creative expression, and aesthetic enjoyment. We also believe the Where’s Wally is the important design features for MOOC because the fact that a very high portion of all sensory learning is visual. From instructor’s perspective, the implication is linking the problem solving process with Where’s Wally and developing learner’s visual literacy. Instructors can utilize this visually engaging mechanics to engage learners in the course. The practical implication for MOOC website designer is to design various faintly discernible cues (e.g., the sparkling star or exclamation mark) and display different cues according to learner’s skill level. System can also give obvious cues as learners are stuck.

Finally, trophies and badges, the fifth most engaging mechanic, mean that learners are engaged by collecting the trophies and badges provided by MOOCs. Most learners prefer to collect as more types of trophies and badges as possible. This finding is consistent with Law and his colleagues (2011). They examined the relationship between gamification and the sustainability of mobile learning applications, and proposed that a badge collection is a crucial enhancer of users’ engagement. Learners who enjoy collecting various types of badges are likely to engage in using mobile applications. From instructor’s perspective, the implication is immediately reward learner’s achievement using trophies and badges and develop milestone badges to enhance learner’s motivation by collecting the badges. The practical implication for MOOC website designer is to design trophies and badge ladder and describe the particularity of them. Moreover, the designer should display the trophies and badges information inside the learner’s personal page and information portion of the comment page.

**Conclusion and suggestions**

In this study, we identified engaging gamification mechanics that influence learners’ engagement in MOOCs. We proposed an empirical approach to identify 40 engaging gamification mechanics for MOOCs, among which the mechanic of the Where’s Wally game has not been revealed in previous gamification studies. A reasonable explanation might be that learners tend to try something challenging and become immersed in the problem-solving process.

This research has both theoretical and practical contributions. From a theoretical standpoint, although previous studies have mentioned certain gamification design factors, they have not provided a conceptual framework based on a theoretical foundation. Therefore, they have not covered engaging gamification factors comprehensively or identified unnecessary factors. Most studies have failed to provide empirical validation of the gamification factors they have discussed. To solve these problems, we constructed a hierarchical framework of gamification and systematically validated the engaging mechanics.

Previous studies have not established a relation between conceptual factors and concrete gamification mechanics. Therefore, even if MOOC operators know which factors are engaging, they do not know how to implement the concepts into practical system mechanics. This study presents a systematic framework of gamification factors and mechanics, which can assist MOOC operators to improve their users’ engagement. Moreover, this paper also provides MOOC operators with empirical data that show which gamification mechanics warrant investigation.

Future research efforts may focus on the connection between the use of gamification mechanics and learning outcomes, since greater numbers of gamification mechanics do not necessary guarantee better learning performance. The limitations of this research should be noted. We do not suggest that the explored gamification mechanics we have discussed represent an exhaustive list. Future research can use various methodologies, such as longitudinal studies, focus groups, and the ethnography approach, to identify other potential gamification mechanics for MOOCs.
References


Choi, D., & Kim, J. (2004). Why people continue to play online games: In search of critical design factors to increase customer loyalty to online contents. *Cyber Psychology & Behavior, 7*(1), 11-24.


Appendix 1. Information of focus group interview

1. Focus group semi-structured questions

   **Learner-learner interaction**
   (1). What learner-learner interaction that you have experienced with the use of MOOCs is the most important in your learning?
   (2). What are the engaging gamification designs for learner-learner interaction that you have experienced with the use of MOOCs? Do you have any example?
   (3). What further gamification designs for learner-learner interaction in MOOC is required?

   **Learner-content interaction**
   (4). What learner-content interaction that you have experienced with the use of MOOCs is the most important in your learning?
   (5). What are the engaging gamification designs for learner-content interaction that you have experienced with the use of MOOCs? Do you have any example?
   (6). What further gamification designs for learner-content interaction in MOOC is required?

   **Learner-instructor interaction**
   (7). What learner-instructor interaction that you have experienced with the use of MOOCs is the most important in your learning?
   (8). What are the engaging gamification designs for learner-instructor interaction that you have experienced with the use of MOOCs? Do you have any example?
   (9). What further gamification designs for learner-instructor interaction in MOOC is required?

2. The record of Focus group interview process

   ![Figure 1. The picture of focus group interview](image)
Appendix 2. The definition of gamification mechanics in MOOCs

<table>
<thead>
<tr>
<th>Types of gamification factors</th>
<th>Gamification mechanics</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learner–content interaction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-expression</td>
<td>GM1: Virtual goods</td>
<td>Virtual goods are nonphysical objects and money purchased for use in online communities or online games (e.g., new avatar options, clothes, weapons, and items).</td>
</tr>
<tr>
<td></td>
<td>GM2: Personal spaces</td>
<td>Personal space is the region surrounding a person that they regard as psychologically theirs.</td>
</tr>
<tr>
<td></td>
<td>GM3: Avatars</td>
<td>Avatars refer to the virtual personality connected with the screen name or handle of an Internet learner.</td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>GM4: Revealing, hiding, and combining items</td>
<td>Revealing, hiding, and combining items are web pages containing learning items that enable learners to explore and reorganize.</td>
</tr>
<tr>
<td></td>
<td>GM5: Where's Wally game</td>
<td>Learners are challenged to find a specific learning object hidden in the website.</td>
</tr>
</tbody>
</table>
| | GM6: Memory-game interactions | Memory-game interactions are game interactions in which a set of cards are laid face down on a surface and two objects are...
flipped face up after each turn. The objective of the game is to turn over pairs of matching objects.

<table>
<thead>
<tr>
<th>Time pressure</th>
<th>GM7: Animated countdown timer</th>
<th>An animated countdown timer is a interface design that counts down in seconds, minutes, hours, and days to any date.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GM8: Time bar</td>
<td>A time bar is a graphical representation that shows the beginning, the duration, and the end of the learning course in MOOCs.</td>
</tr>
<tr>
<td></td>
<td>GM9: Check points</td>
<td>Learners can receive immediate feedback as they progress in the learning tasks.</td>
</tr>
<tr>
<td>Status</td>
<td>GM10: Experience points</td>
<td>Learners can accumulate quantitative data to demonstrate their mastery of skills or knowledge.</td>
</tr>
<tr>
<td></td>
<td>GM11: VIP</td>
<td>VIPs are specific learners who receive high attention as they log on MOOCs.</td>
</tr>
<tr>
<td></td>
<td>GM12: Level-up</td>
<td>Level-up gives learner’s notifications as they achieve certain skill or knowledge level.</td>
</tr>
<tr>
<td>Learner–instructor interaction</td>
<td>GM13: Trophies and badges</td>
<td>Trophies and badges are the most common recognition items found in games because of their versatility and flexibility.</td>
</tr>
<tr>
<td></td>
<td>GM14: Progress bars</td>
<td>Progress bar is a visualized representation can be used to show a user how far along he/she is in a process.</td>
</tr>
<tr>
<td></td>
<td>GM15: Percentage</td>
<td>Percentage shows the progress ratio of learners who have completed a specific goal.</td>
</tr>
<tr>
<td>Instruction</td>
<td>GM16: Personate helper</td>
<td>Personate helper is an anthropomorphic robot that guides new learners to become familiar with the system.</td>
</tr>
<tr>
<td></td>
<td>GM17: Cartoon guide</td>
<td>Cartoon guides are characters that instruct learners as they enter the system.</td>
</tr>
<tr>
<td></td>
<td>GM18: Billboard</td>
<td>Billboards are guidelines that guide learner by providing instructional information.</td>
</tr>
<tr>
<td>Rewards</td>
<td>GM19: Virtual gifts</td>
<td>Virtual gifts are nonphysical objects that work as positive reinforcer.</td>
</tr>
<tr>
<td></td>
<td>GM20: Virtual currency</td>
<td>Virtual currency is a type of unregulated, digital money that is issued and typically controlled by its developers, and used and accepted among the members of a specific virtual community.</td>
</tr>
<tr>
<td></td>
<td>GM21: Praise</td>
<td>Praise is the act of making positive statements about a person, object, or idea, either in public or privately.</td>
</tr>
<tr>
<td></td>
<td>GM22: Recognition</td>
<td>Learners can receive recognition from others based on their excellent learning performance.</td>
</tr>
<tr>
<td>Learner–learner interaction</td>
<td>GM23: Redeemable points</td>
<td>Redeemable points are what learners earn and use to redeem virtual items/ rewards.</td>
</tr>
<tr>
<td></td>
<td>GM24: Skill points</td>
<td>Skill points are assigned for specific activities within the game and are tangential to both XP and RP. They are a bonus set of points that allow players to gain experience or rewards for activities alongside the core.</td>
</tr>
<tr>
<td>GM25: Talent trees</td>
<td>Talent trees are one of the categories in which a learner’s talents are divided. It is so named because the talents branch out, similar to a tree structure.</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>GM26: Karma points</td>
<td>Karma points do not allow players to gain benefit from keeping their karma points, only from sharing them. Karma points are frequently given as part of a regular grind, or check-in behavior, for example: 3 karma points are earned for each monthly check in.</td>
<td></td>
</tr>
<tr>
<td>Peer tutoring</td>
<td>Peer-led team learning is a model of teaching undergraduate science, math, and engineering courses that introduces peer-led workshops as an integral part of a course. Students who have performed strongly in a course are recruited to become peer leaders.</td>
<td></td>
</tr>
<tr>
<td>GM28: Peer mentoring</td>
<td>Peer mentoring is a form of mentorship that typically occurs between a person who has lived through a specific experience (peer mentor) and a person who is new to that experience (the peer mentee).</td>
<td></td>
</tr>
<tr>
<td>GM29: Learning forums</td>
<td>Learning forums are platforms that enable learners to share learning tips, experience, resources, and knowledge.</td>
<td></td>
</tr>
<tr>
<td>Competition</td>
<td>GM30: Hall of fame</td>
<td>A hall of fame is a structure housing memorials to famous or illustrious learners.</td>
</tr>
<tr>
<td>GM31: Team leaderboards</td>
<td>Team leaderboards list winners’ teams and encourage competitions among all learning groups.</td>
<td></td>
</tr>
<tr>
<td>GM32: Individual leaderboards</td>
<td>Individual leaderboards rank MOOC learners according to their learning achievement and list individual winners among all competitors.</td>
<td></td>
</tr>
<tr>
<td>GM33: Gift giving</td>
<td>Gift giving is an expression of love or friendship by giving virtual gifts/items to other learners.</td>
<td></td>
</tr>
<tr>
<td>Altruism</td>
<td>GM34: Gift recommendation</td>
<td>Gift recommendation provides learners with appropriate gift suggestions for social interaction purposes.</td>
</tr>
<tr>
<td>GM35: Gift reminders</td>
<td>Gift reminders is used for calling attention to the gift-giving behavior of learners.</td>
<td></td>
</tr>
<tr>
<td>Group identification</td>
<td>GM36: Self-organized quest teams</td>
<td>Self-organized quest teams allow learners to organize quest teams to solve complex quests or questions.</td>
</tr>
<tr>
<td>GM37: List of family members</td>
<td>Lists of family members allow learners to organize and join learning community to interact and learn together.</td>
<td></td>
</tr>
<tr>
<td>Peer appraisal</td>
<td>GM38: Peer grading</td>
<td>Peer grading means a peer arrangement in which people consider the amount, level, value, worth, quality, or success of the products or outcomes of learning peers of similar status.</td>
</tr>
</tbody>
</table>
Appendix 3. FUZZY AHP PAIRWISE QUESTIONNAIRE

1. Introduction:

Please compare in pairs the relative engagingness between two given item statements regarding the engaging gamification mechanics. If a criterion (or sub-criterion) on the left is more engaging than the one matching on the right, put your check mark to the left of the engagingness “Equal” under the engaging level you prefer. If a criterion (or sub-criterion) on the left is less engaging than the one matching on the right, put your check mark to the right of the engagingness “Equal” under the engaging level you prefer. The notations of relative engagingness are following:

1. Absolutely – Absolutely more engaging
2. Very strongly – Very strongly more engaging
3. Strongly – Strongly more engaging
4. Weakly – Weakly more engaging
5. Equally – Equally engaging

2. The definition of gamification:

Gamification means game mechanics into nongame settings to increase user engagement and enjoyment of a product or service, and to encourage users to perform certain behaviors. Gamification essentially functions as entertainment that causes learners to enjoy actively participating and engaging with others, such as through reputation points, rewards, and goal setting.

3. The definition of gamification mechanics in Massive open online courses (MOOCs)

<table>
<thead>
<tr>
<th>Types of gamification</th>
<th>Gamification factors</th>
<th>Gamification mechanics</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner-content interaction</td>
<td>Self-expression</td>
<td>GM1: Virtual goods</td>
<td>Virtual goods are nonphysical objects and money purchased for use in online communities or online games (e.g., new avatar options, clothes, weapons, and items).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM2: Personal spaces</td>
<td>Personal space is the region surrounding a person that they regard as psychologically theirs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM3: Avatars</td>
<td>Avatars refer to the virtual personality connected with the screen name or handle of an Internet learner.</td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>Revealing, hiding, and combining items</td>
<td>GM4: Revealing, hiding, and combining items</td>
<td>Revealing, hiding, and combining items are web pages contains learning items that enable learners to explore and reorganize.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM5: Where's Wally game</td>
<td>Learners are challenged to find a specific learning object hidden in the website.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM6: Memory-game interactions</td>
<td>Memory-game interactions are game interactions in which a set of cards are laid face down on a surface and two objects are flipped face up after each turn. The objective of the game is to turn over pairs of matching objects.</td>
</tr>
<tr>
<td>Time pressure</td>
<td>GM7: Animated countdown timer</td>
<td></td>
<td>An animated countdown timer is a interface design that counts down in seconds, minutes, hours, and</td>
</tr>
</tbody>
</table>
days to any date.

GM8: Time bar
A time bar is a graphical representation that shows the beginning, the duration, and the end of the learning course in MOOCs.

GM9: Check points
Learners can receive immediate feedback as they progress in the learning tasks.

GM10: Experience points
Learners can accumulate quantitative data to demonstrate their mastery of skills or knowledge.

GM11: VIP
VIPs are specific learners who receive high attention as they log on MOOCs.

GM12: Level-up
Level-up gives learner’s notifications as they achieve certain skill or knowledge level.

GM13: Trophies and badges
Trophies and badges are the most common recognition items found in games because of their versatility and flexibility.

GM14: Progress bars
Progress bar is a visualized representation can be used to show a user how far along he/she is in a process.

GM15: Percentage
Percentage shows the progress ratio of learners who have completed a specific goal.

GM16: Personate helper
Personate helper is an anthropomorphic robot that guides new learners to become familiar with the system.

GM17: Cartoon guide
Cartoon guides are characters that instruct learners as they enter the system.

GM18: Billboard
Billboards are guidelines that guide learner by providing instructional information.

GM19: Virtual gifts
Virtual gifts are nonphysical objects that work as positive reinforcer.

GM20: Virtual currency
Virtual currency is a type of unregulated, digital money that is issued and typically controlled by its developers, and used and accepted among the members of a specific virtual community.

GM21: Praise
Praise is the act of making positive statements about a person, object, or idea, either in public or privately.

GM22: Recognition
Learners can receive recognition from others based on their excellent learning performance.

GM23: Redeemable points
Redeemable points are what learners earn and use to redeem virtual items/rewards.

GM24: Skill points
Skill points are assigned for specific activities within the game and are tangential to both XP and RP. They are a bonus set of points that allow players to gain experience or rewards for activities alongside the core.

GM25: Talent trees
Talent trees are one of the categories in which a learner’s talents are divided. It is so named because the talents branch out, similar to a tree structure.

GM26: Karma points
Karma points do not allow players to gain benefit from keeping their karma points, only from sharing them. Karma points are frequently given as part of a regular grind, or check-in behavior, for example: 3
Peer tutoring  
GM27: Peer-led team learning  
Peer-led team learning is a model of teaching undergraduate science, math, and engineering courses that introduces peer-led workshops as an integral part of a course. Students who have performed strongly in a course are recruited to become peer leaders.

GM28: Peer mentoring  
Peer mentoring is a form of mentorship that typically occurs between a person who has lived through a specific experience (peer mentor) and a person who is new to that experience (the peer mentee).

GM29: Learning forums  
Learning forums are platforms that enable learners to share learning tips, experience, resources, and knowledge.

Competition  
GM30: Hall of fame  
A hall of fame is a structure housing memorials to famous or illustrious learners.

GM31: Team leaderboards  
Team leaderboards list winners’ teams and encourage competitions among all learning groups.

GM32: Individual leaderboards  
Individual leaderboards rank MOOC learners according to their learning achievement and list individual winners among all competitors.

Altruism  
GM33: Gift giving  
Gift giving is an expression of love or friendship by giving virtual gifts/ items to other learners.

GM34: Gift recommendation  
Gift recommendation provides learners with appropriate gift suggestions for social interaction purposes.

GM35: Gift reminders  
Gift reminders are used for calling attention to the gift-giving behavior of learners.

Group identification  
GM36: Self-organized quest teams  
Self-organized quest teams allow learners to organize quest teams to solve complex quests or questions.

GM37: List of family members  
Lists of family members allow learners to organize and join learning community to interact and learn together.

Peer appraisal  
GM38: Peer grading  
Peer grading means a peer arrangement in which people consider the amount, level, value, worth, quality, or success of the products or outcomes of learning peers of similar status.

GM39: Peer comments  
Peer comments are the constructive feedback provided by other learning community members.

GM40: Peer emoticon feedback  
Peer emoticon feedback refers to a meta-communicative pictorial representation of a facial expression that allows learners to interact with each other.
### Part 1: Three gamification dimensions

<table>
<thead>
<tr>
<th>Learner–content interaction</th>
<th>Learner–content interaction</th>
<th>Learner–instructor interaction</th>
<th>Learner–instructor interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolutely</td>
<td>9:1</td>
<td>8:1</td>
<td>7:1</td>
</tr>
<tr>
<td>Very strongly</td>
<td>7:1</td>
<td>6:1</td>
<td>5:1</td>
</tr>
<tr>
<td>Strongly</td>
<td>5:1</td>
<td>4:1</td>
<td>3:1</td>
</tr>
<tr>
<td>Weakly</td>
<td>3:1</td>
<td>2:1</td>
<td>1:1</td>
</tr>
<tr>
<td>Equally</td>
<td>1:1</td>
<td>1:2</td>
<td>1:3</td>
</tr>
<tr>
<td>Weakly</td>
<td>1:4</td>
<td>1:5</td>
<td>1:6</td>
</tr>
<tr>
<td>Strongly</td>
<td>1:7</td>
<td>1:8</td>
<td>1:9</td>
</tr>
</tbody>
</table>

### Part 2: Thirteen gamification factors under dimensions

<table>
<thead>
<tr>
<th>Self-expression interaction</th>
<th>Pattern recognition interaction</th>
<th>Self-expression time pressure interaction</th>
<th>Status recognition interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal setting</td>
<td>Instruction</td>
<td>Rewards</td>
<td>Peer tutoring</td>
</tr>
<tr>
<td>Reputation points</td>
<td>Competition</td>
<td>Altruism</td>
<td>Group identification</td>
</tr>
<tr>
<td>Reputation points</td>
<td></td>
<td></td>
<td>Peer appraisal</td>
</tr>
</tbody>
</table>

| Reputation points           |                                  |                                          | Competition                   |
| Peer tutoring               | Altruism                        | Group identification                     | Peer appraisal                |
| Peer tutoring               |                                  |                                          | Altruism                      |
| Peer tutoring               |                                  |                                          | Group identification          |

197
### Part 3: Forty gamification mechanics under factors

<table>
<thead>
<tr>
<th>Criterion (or sub-criterion)</th>
<th>Absolutely</th>
<th>Very strongly</th>
<th>Strongly</th>
<th>Weakly</th>
<th>Equally</th>
<th>Weakly</th>
<th>Strongly</th>
<th>Very strongly</th>
<th>Absolutely</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM1: Virtual goods</td>
<td>9:1</td>
<td>8:1</td>
<td>7:1</td>
<td>6:1</td>
<td>5:1</td>
<td>4:1</td>
<td>3:1</td>
<td>2:1</td>
<td></td>
</tr>
<tr>
<td>GM1: Virtual goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM2: Personal spaces</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM4: Revealing, hiding, and combining items</td>
<td>1:1</td>
<td>1:2</td>
<td>1:3</td>
<td>1:4</td>
<td>1:5</td>
<td>1:6</td>
<td>1:7</td>
<td>1:8</td>
<td>1:9</td>
</tr>
<tr>
<td>GM5: Where's Wally game</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM7: Animated countdown timer</td>
<td>9:1</td>
<td>8:1</td>
<td>7:1</td>
<td>6:1</td>
<td>5:1</td>
<td>4:1</td>
<td>3:1</td>
<td>2:1</td>
<td></td>
</tr>
<tr>
<td>GM8: Time bar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM10: Experience points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM11: VIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM12: Level-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM13: Trophies and badges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM14: Progress bars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM15: Percentage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM16: Personate helper</td>
<td>GM17: Cartoon guide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM18: Billboard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM19: Virtual gifts</td>
<td>GM20: Virtual currency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM21: Praise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM22: Recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM23: Virtual gifts</td>
<td>GM24: Skill points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM25: Talent points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM26: Karma points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM27: Peer-led team learning</td>
<td>GM28: Peer mentoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM29: Learning forums</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM30: Hall of fame</td>
<td>GM31: Team leaderboards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM32: Individual leaderboards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM33: Gift giving</td>
<td>GM34: Gift recommendation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM35: Gift reminders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM36: Self-organized quest teams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM37: List of family members</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM38: Peer grading</td>
<td>GM39: Peer comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GM40: Peer emoticon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4. A fuzzy-AHP approach for determining relative engagingness of gamification mechanics in MOOCs

Step 1. Constructing the fuzzy comparison matrix

Triangular fuzzy numbers $\tilde{M}_{ij}$ from $1$ to $9$ was employed to represent the results of users’ assessments of the pairwise comparisons between each of the gamification mechanics by constructing a fuzzy positive reciprocal matrix $M$. The proposed fuzzy comparison matrix was defined as follows:

$$M = \begin{bmatrix} \tilde{M}_{ij} \end{bmatrix}$$

$M$ : fuzzy positive reciprocal matrix

$$\tilde{M}_{ij} = (L_{ij}, M_{ij}, R_{ij})$$

$L_{ij}$ : the left value of the fuzzy membership function of the collected subject assessments of design mechanic $j$ of decision element $i$

$M_{ij}$ : the middle value of the fuzzy membership function of the collected subject assessments of design mechanic $j$ of decision element $i$

$R_{ij}$ : the right value of the fuzzy membership function of the collected subject assessments of design mechanic $j$ of decision element $i$

$$\tilde{M}_{ij} = \frac{1}{M_{ji}}, \forall i, j = 1, 2, \ldots, n$$

Step 2. Integration of the collected subjects’ assessments of each decision element

There are many possible approaches to integrating subject assessments when calculating the triangular fuzzy number. In contrast to some studies that apply statistical parameters such as the minimum, maximum, mean $\alpha$ and mode to represent the fuzzy numbers, this study applied the geometric mean method proposed by Buckley [59]. The computing process is defined as follows:

$$\tilde{m}_{ij} = \left( \frac{1}{n} \right) \bigotimes \left( \tilde{m}_{ij}^1 \oplus \tilde{m}_{ij}^2 \oplus \cdots \oplus \tilde{m}_{ij}^n \right)$$

$\tilde{m}_{ij}$ : Integrated triangular fuzzy numbers

$\tilde{m}_{ij}^N$ : The value of the pair comparison of the collected subject assessments of design mechanic $j$ of decision factor $i$

$n$ : The number of subjects

Step 3. Computation of fuzzy weight

After integrating the collected data and calculating the corresponding triangular fuzzy numbers, we used the Approximation Method proposed by Buckley [59] to compute the fuzzy weight. The formula of the Approximation Method for computing the fuzzy weights is defined as follows:

$$\tilde{Z}_i = \left( a_{i1} \otimes a_{i2} \otimes \cdots \otimes a_{in} \right)^{1/n}, \forall i = 1, 2, \ldots, n$$

200
\[ \tilde{W}_i = \tilde{Z}_i \otimes \left( \tilde{Z}_1 \oplus \tilde{Z}_2 \oplus \cdots \oplus \tilde{Z}_n \right)^{-1} \]

\( \tilde{Z}_i \): The geometric mean value of the triangular fuzzy number

\( \tilde{a}_{ij} \): The triangular fuzzy number of row \( i \) and column \( j \) in the fuzzy positive reciprocal matrix

\( \tilde{W}_i \): The fuzzy weight of each row of the fuzzy positive reciprocal matrix

**Step 4. Defuzzification of decision elements**

The weights of the decision elements were represented by fuzzy values. The defuzzification process assigned a distinct number to each of the decision element. We then used the Center of Gravity Method of defuzzification to calculate the center of gravity of the triangular fuzzy number. Given a triangular fuzzy number and its three sides, denoted by \( \tilde{A} = (L_{ij}, M_{ij}, R_{ij}) \), the defuzzified weight \( DF_{ij} \) was calculated using the following formula:

\[ DF_{ij} = \frac{1}{3} \left( R_{ij} - L_{ij} \right) + \frac{1}{3} \left( M_{ij} - L_{ij} \right) / 3 + L_{ij} \]

**Step 5. Normalization of defuzzified weights**

To compare the relative engagingness of different decision element at different levels, we first normalized the defuzzified weights. The definition of the normalized weights \( NW_{ij} \) of each decision dimension at each level can be defined as follows:

\[ NW_{ij} = DF_{ij} / \sum DF_{ij} \]

**Step 6. Calculation of the synthesized weight for each element at each level**

We calculated the normalized weights of each element at each level after step 5. However, to determine the priority of each mechanic, it was still necessary to synthesize weights for each decision element at each decision level. The larger the value of the synthesized weight, the higher the priority of the dimension. The definition of synthesized weights of each decision element at each level was defined as follows:

\[ NW_{ij} = NW_{ij} \times NW_{ij} \times NW_{ij} \]

**Step 7. Checking for consistency**

Consistency Index (CI) was employed to designate overall inconsistency for the proposed hierarchy and for each decision dimension. Consistency Ratio (CR) was also calculated to describe the consistency of the pair-wise comparisons. The equations for calculating CI and CR for each decision were:

\[ \text{Consistency Index (CI)} = \frac{\lambda_{\text{Max}} - n}{n - 1} \]

where \( \lambda_{\text{Max}} \) is the maximum eigenvalue, and \( n \) the number of decision component

\[ \text{Consistency Ratio (CR)} = \frac{\text{CI}}{RI} \]

\( RI \) is the average index for randomly generated weights obtained from a table of random consistency indices. To judge the consistency of the pair-wise outputs, if \( CR \) was \( \leq 0.1 \), then the output of the pair-wise comparison was sufficiently consistent. On the other hand, if \( CR \) was \( > 0.1 \), then the results of the pair-wise comparison were inconsistent.
## Appendix 5. The Fuzzy AHP Weight of Gamification factors and mechanics table

<table>
<thead>
<tr>
<th>Criteria (Layer 1)</th>
<th>Weights (Layer 2)</th>
<th>Sub-criteria (Layer 3)</th>
<th>Local weights</th>
<th>Sub-criteria (Layer 3)</th>
<th>Local weights</th>
<th>Global weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner–content interaction</td>
<td>0.281</td>
<td>Self-expression</td>
<td>0.372</td>
<td>GM1: Virtual gifts</td>
<td>0.911</td>
<td>0.09520</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM2: Personal spaces</td>
<td>0.032</td>
<td>0.00335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM3: Avatars</td>
<td>0.057</td>
<td>0.00599</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pattern recognition</td>
<td>0.420</td>
<td>GM4: Revealing, hiding, and combining items</td>
<td>0.279</td>
<td>0.02293</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM5: Where's Wally game</td>
<td>0.403</td>
<td>0.04760</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM6: Memory-game interactions</td>
<td>0.318</td>
<td>0.03310</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time pressure</td>
<td>0.171</td>
<td>GM7: Animated countdown timer</td>
<td>0.196</td>
<td>0.00942</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM8: Time bar</td>
<td>0.203</td>
<td>0.00975</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM9: Check points</td>
<td>0.601</td>
<td>0.02890</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Status</td>
<td>0.037</td>
<td>GM10: Experience points</td>
<td>0.332</td>
<td>0.00345</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM11: VIP</td>
<td>0.312</td>
<td>0.00324</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM12: Level-up</td>
<td>0.356</td>
<td>0.00370</td>
</tr>
<tr>
<td>Learner–instructor interaction</td>
<td>0.316</td>
<td>Goal setting</td>
<td>0.313</td>
<td>GM13: Trophies and badges</td>
<td>0.466</td>
<td>0.04610</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM14: Progress bars</td>
<td>0.260</td>
<td>0.02572</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM15: Percentage</td>
<td>0.274</td>
<td>0.02709</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instruction</td>
<td>0.259</td>
<td>GM16: Personate helper</td>
<td>0.334</td>
<td>0.02734</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM17: Cartoon guide</td>
<td>0.333</td>
<td>0.02725</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM18: Billboard</td>
<td>0.333</td>
<td>0.02725</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rewards</td>
<td>0.428</td>
<td>GM19: Virtual gifts</td>
<td>0.253</td>
<td>0.02422</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM20: Virtual currency</td>
<td>0.251</td>
<td>0.02395</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM21: Praise</td>
<td>0.271</td>
<td>0.02665</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM22: Recognition</td>
<td>0.225</td>
<td>0.02043</td>
</tr>
<tr>
<td>Learner–learner interaction</td>
<td>0.403</td>
<td>Reputation points</td>
<td>0.410</td>
<td>GM23: Redeemable points</td>
<td>0.511</td>
<td>0.08450</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM24: Skill points</td>
<td>0.166</td>
<td>0.02740</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM25: Talent trees</td>
<td>0.160</td>
<td>0.02647</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM26: Karma points</td>
<td>0.163</td>
<td>0.02687</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peer tutoring</td>
<td>0.052</td>
<td>GM27: Peer-led team learning</td>
<td>0.311</td>
<td>0.00652</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM28: Peer mentoring</td>
<td>0.311</td>
<td>0.00652</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM29: Learning forums</td>
<td>0.378</td>
<td>0.00791</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Competition</td>
<td>0.201</td>
<td>GM30: Hall of fame</td>
<td>0.042</td>
<td>0.00340</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM31: Team leaderboards</td>
<td>0.906</td>
<td>0.07340</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GM32: Individual leaderboards</td>
<td>0.052</td>
<td>0.00420</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Altruism</td>
<td>0.039</td>
<td>GM33: Gift giving</td>
<td>0.423</td>
<td>0.00665</td>
</tr>
<tr>
<td>Group identification</td>
<td>0.043</td>
<td>GM34: Gift recommendation</td>
<td>0.374</td>
<td>0.00587</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>---------------------------</td>
<td>-------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM35: Gift reminders</td>
<td>0.203</td>
<td>0.00320</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM36: Self-organized quest teams</td>
<td>0.613</td>
<td>0.01063</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM37: List of family members</td>
<td>0.387</td>
<td>0.00670</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer appraisal</td>
<td>0.255</td>
<td>GM38: Peer grading</td>
<td>0.404</td>
<td>0.04150</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM39: Peer comments</td>
<td>0.214</td>
<td>0.02200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM40: Peer emoticon feedback</td>
<td>0.382</td>
<td>0.03930</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Subgroup Discovery with User Interaction Data: An Empirically Guided Approach to Improving Intelligent Tutoring Systems

Eric G. Poitras*, Susanne P. Lajoie, Tenzin Doleck and Amanda Jarrell

1University of Utah, USA // 2McGill University, Canada // eric.poitras@utah.edu // susanne.lajoie@mcgill.ca // tenzin.doleck@mail.mcgill.ca // amanda.jarrell@mail.mcgill.ca

*Corresponding author

(Submitted November 24, 2014; Revised May 12, 2015; Accepted July 31, 2015)

ABSTRACT

Learner modeling, a challenging and complex endeavor, is an important and oft-studied research theme in computer-supported education. From this perspective, Educational Data Mining (EDM) research has focused on modeling and comprehending various dimensions of learning in computer based learning environments (CBLE). Researchers and designers are actively attempting to improve learning systems by incorporating adaptive mechanisms that respond to the varying needs of learners. Recent advances in data mining techniques provide new possibilities and exciting opportunities for developing adaptive systems to better support learners. This study is situated in the context of clinical reasoning in an Intelligent Tutoring System called BioWorld and it aims to examine the relationship between the lab-tests ordered and misconceptions held by learners. Toward this end, we employ an EDM technique called subgroup discovery to unpack the rules that embody the hypothesized link. Examining such links may have implications for identifying the points along learning trajectories where learners should be provided the requisite scaffolding. This study represents our efforts to evaluate and derive empirically based design prescriptions for improving Intelligent Tutoring Systems. Implications for practice and future research directions are also discussed.

Keywords

Subgroup discovery, Educational data mining, Intelligent tutoring systems, Medical education, Clinical reasoning, Lab tests, Misconceptions, Learner modeling

Introduction

Learner modeling is a critical component in the process of adapting instruction to the specific needs of different learners with computerized systems. Adaptive instructional systems may be defined as a systematic process which consists of four steps: (1) capturing information about the learner; (2) analyzing learner interactions through a model of learner characteristics in relation to the domain; (3) selecting the appropriate instructional content and resources; and (4) delivering the content to the learner (Shute & Zapata-Rivera, 2012). The analytical function of the learner model can be further classified in terms of processes conducted at both the macro and micro levels (VanLehn, 2006). At the macro-level, a representation of the path towards competency within the domain is updated for each task with the aim of selecting the next task that is the most appropriate for the learner. At the micro-level, instructional materials such as hints and feedback are delivered to the learner on the basis of a representation that is repeatedly updated over the duration of task performance. Intelligent tutoring systems have been shown to improve upon typical classroom instruction in several disciplines, including mathematics (e.g., Cognitive Tutors; Koedinger & Corbett, 2006), computer science (e.g., Constraint-Based Tutors; Mitrovic, 2003), microbiology (e.g., Narrative-Based Tutors; Rowe, Shores, Mott, & Lester, 2011), as well as physics and computer literacy (e.g., Dialogue-Based Tutors; Graesser, VanLehn, Rosé, Jordan, & Harter, 2001).

The challenge in modeling learning in the context of solving ill-structured problems is that there are multiple paths towards attaining the correct solution (Lajoie, 2003, 2009). These paths should be documented in order to represent common misconceptions or impasses in moving along the trajectory towards competency. In the medical domain, experts have been shown to diagnose patient diseases in different ways, while at the same time, justifying plausible hypotheses on the basis of common evidence items, including patient symptoms and lab test information (Gauthier & Lajoie, 2014; Lajoie, Gauthier, & Lu, 2009). An intelligent tutoring system such as BioWorld can represent these evidence items in terms of a novice-expert overlay system, which compares novice solution paths to those of the experts in order to individualize feedback. A similar modeling technique has been applied in other systems, such as SlideTutor in diagnosing dermatopathology (Feyzi-Behnagh, Azevedo, Legowski, Reitmeyer, Tseytlin, & Crowley, 2014) as well as the MedU virtual patient cases (Berman, Fall, Chessman, Dell, Lang, Leong, Nixon, & Smith, 2011). On the basis of the user interactions that are recorded by BioWorld, the system will highlight areas of
similarities and differences with the expert solution path, allowing novices to self-reflect on their own approach to resolving the problem (Lajoie & Poitras, 2014). Several studies investigating the novice-expert overlay model have been carried out with the aim of capturing linguistic features from written case summaries to tailor feedback content (Poitras, Doleck, & Lajoie, 2014; Lajoie, Poitras, Doleck, & Jarrell, 2015) as well as the impact of goal-orientations and affective reactions towards attention given to feedback (Lajoie, Naismith, Poitras, Hong, Panesso-Cruz, Ranelluci, & Wiseman, 2013).

In this study, we examine the use of subgroup discovery for the induction of rules that characterize the relationship between impasses in problem-solving and lab-tests ordered in BioWorld. Whereas our previous research characterized how experts converged in their paths to solving a problem, the present study captures how novices diverged from an expert solution path. In doing so, we claim that subgroup discovery algorithms are particularly well suited towards describing the multiple paths that characterize problem-solving in ill-structured domains. The findings obtained from this analytical approach have implications for evaluating and deriving empirically-based design guidelines for novice-expert overlay models. In particular, the capabilities of this type of system to provide instruction in becoming aware of common impasses and misconceptions in solving certain types of problems, which is prescribed on the basis of the specific needs of different novices. In the following section, we review the theoretical model that accounts for novices’ ability to regulate their efforts in overcoming impasses in problem solving.

**The regulation of learning while solving problems in the medical domain**

Social cognitive models of self-regulation characterize the process of solving ill-structured problems as a recursive and iterative process, where adjustments to the solution are made on the basis of refining the problem space (Zimmerman & Campillo, 2003). As such, self-regulated problem solvers engage in cycles of forethought, performance, and self-reflection (Zimmerman, 2000). In the forethought phase, novices orient themselves in the problem space, at the same time, formulating a plan to solve the problem. The performance phase is characterized by the novices’ efforts to solve the problem by executing the planned steps and monitoring the outcomes. The self-reflection phase involves the novices’ evaluations of the overall progress and elaborations about the problem space, resulting in conclusions about the case. In doing so, the problem-solving process is recursive in that the outcomes of prior steps inform the next ones that are taken to solve the problem.

In accordance with the basic phases of self-regulation, novices in the medical domain regulate their own approaches to solving diagnostic problems in accordance with disciplinary-based practices. Self-regulation involves several types of metacognitive activities; namely, orienting, planning, executing, monitoring, evaluating, and elaborating (Meijer, Veenman, & van Hout-Wolters, 2006; Lu & Lajoie, 2008; Lajoie & Lu, 2012). As an example, self-regulated problem solvers have the ability to notice pertinent vital signs, such as the patient heart rate exceeding the normal range, which could potentially be caused by a tumor of the adrenal glands. To test this assumption the plan might entail testing for pheochromocytoma by ordering a lab test to verify serum levels of the catecholamines adrenalin and noradrenalin. This plan is executed, and the lab test was found to be pertinent, as serum levels were elevated, thereby confirming a diagnosis of pheochromocytoma. Self-regulated problem-solvers evaluate their own progress by re-adjusting the plausibility of differential diagnoses. In doing so, a battery of lab tests might be ordered to rule out commonly known alternative diagnoses. The outcomes of these efforts are evaluated and will inform subsequent attempts to solve the problem as the problem-solver becomes progressively more confident in their own solution.

**Modeling self-regulation in BioWorld**

BioWorld is a computer-based learning environment designed as a cognitive tool (Pea, 1985; Perkins, 1985; Salomon, Perkins, & Globerson, 1991; Derry & Lajoie, 1993; Jonassen & Reeves, 1996; Lajoie, 2000, 2005). Tools embedded in the learning environment aim to support the cognitive and metacognitive activities that mediate performance in diagnosing patient diseases (see Figure 1). In doing so, the system captures learner behaviors that characterize their efforts to regulate several aspects of problem-solving.

Each patient case begins with the case history where novices formulate their differential diagnoses. Once novices select the possible diagnoses that fit with the symptoms, they report their confidence in their primary hypothesis by
using the Belief Meter (% certainty). Novices gather evidence from the case history by highlighting relevant symptoms, the outcome of which is shown in the Evidence Table, which remains visible throughout the problem solving activity. There is a library where novices access additional information about the disease they are investigating. Information in the library represents the typical symptoms and transmission routes of a specific disease, as well as a glossary of medical terminology and diagnostic testing procedures. In order to solve problems, novices order lab tests to confirm or disconfirm specific diagnoses. They do so by ordering tests on the patient chart, where the outcomes of their tests are recorded in the Evidence Table. A consultation tool is present and novices can obtain hints on request, delivered in increasing order of specificity.

A subset of user interactions with interface elements of BioWorld are captured and analyzed by the novice-expert overlay model. These interactions include the evidence items posted to the evidence palette, namely, information regarding patient symptoms and lab-tests ordered in the chart. The user must identify a particular item as pertinent to solving the case by sending it to the evidence palette. The evidence palette thus serves as a monitoring tool, allowing the user to review the evidence items, otherwise referred to as the steps taken to solve the problem. Novices and experts have been found to interact differently with this tool, as monitoring is likely to develop along a trajectory towards becoming more proficient in the domain (Lajoie et al., 2013).

Figure 1. A screenshot of the BioWorld interface

Once novices have submitted their final diagnosis, the novice-expert overlay model individualizes the feedback that is delivered to novices after the submission of their final solution for a particular case. Before novices attend to the feedback, they are required to categorize the evidence items that either confirms, refutes, or that are irrelevant to their final diagnosis. Novices must then prioritize the evidence items in terms of their relative importance in solving the case. In order to foster novices’ self-reflection, BioWorld provides them with a case summary and a student report. The case summary consists of a writing activity where novices build a justification for their final diagnosis on the basis of the steps taken to solve the case. The student report refers to the formative feedback that is provided to the novices by the system, where the solution steps of a validated expert solution path is compared to the novices’ solution in order to highlight similarities and differences. The expert solution path also provides a case summary,
written by the expert, which outlines in detail the steps that were taken to solve the case, and how each step contributed to formulating the final diagnosis.

**Improving the novice-expert overlay model**

This study aims to further refine the capabilities of the novice-expert overlay model to individualize instruction by mining learner behaviors to uncover common patterns that are indicative of impasses in problem-solving. Does the application of subgroup discovery to uncover relationships between errors in diagnostic reasoning and the nature of the lab tests that are ordered by novices lead to improvements in the design guidelines of the novice-expert overlay model? We focus on two research questions: (a) What type of diagnostic tests ordered by novices are most often found to precede diagnostic errors during problem-solving? (b) What are the impacts of these impasses towards diagnostic outcomes as manifested in problem-solving performance and evaluation?

With regard to the first research question, we hypothesize that impasses in problem-solving, as evidenced by ordering non-pertinent lab-tests, are most often found in solving rare and complex cases in BioWorld. As such, the subgroup discovery algorithm is applied to uncover patterns in logged user interactions while solving three cases in BioWorld with varying levels of difficulty. In order to address the second research question, we hypothesize that common impasses encountered while solving a type of disease are indicative of prominent misconceptions about the nature of the disease. We rely on a combination of measures to assess proficiency in diagnostic reasoning as well as novices understanding of the disease after their attempts to solve the problem in order to ascertain the impacts of these impasses in problem-solving.

**Methods**

**Participants**

The participants consist of 30 undergraduate students who were compensated $20 for practicing diagnostic reasoning with BioWorld during a 2 hour session. The convenience sample includes 11 men (37%) and 19 women (63%) with an average age of 23 (SD = 2.60). Participants were randomly assigned cases to solve, including Amy (Diabetes Mellitus Type 1), Cynthia (Hyperthyroidism), and Susan Taylor (Pheochromocytoma).

**Measures**

*Process measures*

BioWorld logs user interactions while interface elements while novices are solving problems. These logs are stored on a MySQL server database in the form of a timestamp, an identifier (i.e., participant and case ID), a space (i.e., BioWorld interface), the user interaction label (e.g., add test), and description (e.g., Thyroid Stimulating Hormone (TSH) Result: 0.2 mU/L). The learning behaviors that were extracted from the database consists of 172 unique lab tests, ordered by the learners using the chart panel in BioWorld. Each variable is represented by its corresponding label (i.e., add test), the name of the procedure (i.e., Abdominal Exam Result), and the value that was obtained (i.e., Normal). These variables are assigned a value that indicates whether the relevant lab test was ordered or not during a particular line of diagnostic reasoning (i.e., true and false, respectively). As such, BioWorld log files were aggregated at the level of changes in the selected or submitted hypothesis for the purposes of determining the occurrence of each learning behavior within each line in diagnostic reasoning. A total of 304 lines of diagnostic reasoning were aggregated by following this procedure.

*Product measures*

On the basis of the logged user interactions, several performance metrics can be calculated in order to appraise learner performance in solving problems. We distinguish between three types of performance metrics, namely, diagnostic efficacy (e.g., accuracy, count of matches with experts, and percentage of matches with experts),
efficiency (e.g., number of tests ordered and time to solve the case), and affect (e.g., confidence). The target variable that was extracted from the database consists of diagnostic accuracy (i.e., Correct, Incorrect) in order to discover interesting subgroups for the incorrect diagnoses that are indicative of common misconceptions or impasses in solving a case. This target attribute was evenly distributed, with frequencies of 156 correct and 148 incorrect lines of diagnostic reasoning. The accuracy of the diagnosis was determined at the moment when either the learner selected a main hypothesis or the final hypothesis was submitted as the solution for the case. BioWorld logs record changes to the selected or submitted hypotheses, allowing the system to track distinct lines in diagnostic reasoning that are indicative of instances when learners reach an impasse.

Experimental procedure

The data was collected as part of an experiment that examined the antecedent factors to attention allocated towards feedback in BioWorld (see Naismith, 2013). First, participants were asked to complete a demographic questionnaire and the achievement goal questionnaire (i.e., PALS; Midgley, Maehr, Hruda, Anderman, Anderman, Freeman et al., 2000). A training session enabled the participants to learn how to use BioWorld by solving a practice case that was unrelated to the actual cases solved during the study. During the study, the participants performed a think aloud protocol, which involved participants verbalizing their own thought processes while solving a case (see Ericsson & Simon, 1993). At the end of each case, participants filled out the feedback emotions questionnaire (i.e., AEQ; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). The order of cases to be solved was counterbalanced to mitigate practice effects. The average length of the study was two hours to solve a total of three cases.

Subgroup discovery algorithm

Subgroup discovery (Wrobel, 1997; Klösgen, 2002) is an educational data mining technique that aims to discover interesting rules (i.e., as generalizable as possible and with the most unusual statistical characteristics) with respect to a set of learning behaviors. The technique involves searching for relationships between these different learning variables and a target variable, which are described in terms of individual rules or subgroups. A rule is defined as a conjunction of learning behaviors (i.e., learning behavior-value pair) that is able to account for an unusual statistical distribution in relation to the variable of interest (i.e., target variable). The main difference between the subgroup discovery and the classification task is that subgroup discovery algorithms describe unusual relations between learning behaviors and a certain value of the target variable in a comprehensive manner. By contrast, classification algorithms predict these relationships with an emphasis on precision and interpretability (see Herrera, Carmona, González, & del Jesus, 2011).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Definition</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive instances</td>
<td>Count of examples covered by rule and correctly included</td>
<td>N(Antecedent = True &amp; Conclusion = True)</td>
</tr>
<tr>
<td>Negative instances</td>
<td>Count of examples covered by rule and incorrectly included</td>
<td>N(Antecedent = True &amp; Conclusion = False)</td>
</tr>
<tr>
<td>Size</td>
<td>Count of examples covered by rule</td>
<td>N(Antecedent = True)</td>
</tr>
<tr>
<td>Coverage</td>
<td>Percentage of size to the total number of examples</td>
<td>N(Antecedent = True &amp; Conclusion = True)/N(Examples)</td>
</tr>
<tr>
<td>Complexity length</td>
<td>Count of variables in antecedent</td>
<td>N(Variables in Antecedent)</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>Percentage of correct inclusions to the count of examples covered by the rule</td>
<td>N(Antecedent = True &amp; Conclusion = True)/N(Antecedent = True)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Percentage of correct inclusions and exclusions to the total number of examples</td>
<td>[N(Antecedent = True &amp; Conclusion = True)+N(Antecedent = False &amp; Conclusion = False)]/N(Examples)</td>
</tr>
</tbody>
</table>
The subgroup discovery algorithm performs an exhaustive search of the dataset. The dataset consists of the lines of diagnostic reasoning that correspond to a particular case, where the amount of learning behaviors were reduced by calculating the value of the chi-squared statistic with respect to diagnostic accuracy. Learning behaviors whose weights are greater than a null value were selected for inclusion in the subgroup discovery task. The search task was constrained by a number of parameters to identify antecedents that lead to a conclusion of an incorrect diagnosis. The maximum depth of the search was set to 1 in order to limit the number of task iterations, more specifically the number of antecedents of the generated rules. The minimum coverage was set to 5% to ensure that the misconceptions, although rare in occurrence, are still prevalent enough to warrant intervention.

A number of quality measures are used to determine the interpretability of the rules that are generated by the subgroup discovery task, including measures of generality, complexity, and precision. Generality refers to the extent to which examples are covered by the rule. Complexity consists of the simplicity of a particular rule. Precision is defined as the measure of how well a rule identifies and excludes the examples that belong to a subgroup. Table 1 outlines the metrics associated to these quality measures as well as their definitions and formulas.

The main criterion for the evaluation of the quality of subgroups is precision. This criterion is chosen on the basis of the research objectives as it is the most suitable in terms of identifying patterns of learning behaviors that are indicative of misconceptions in diagnostic reasoning. As such, the minimum value for the quality measure of precision was set to 70% in order to determine rules that perform well in terms of including examples to subgroups of misconceptions.

In the following section, we outline the steps involved in the subgroup discovery task. First, we implemented the subgroup discovery algorithm to extract patterns from lines of diagnostic reasoning that relate lab tests ordered to the correctness of the main hypothesis. Second, we validated subgroups of lines of diagnostic reasoning that are characterized by misconceptions using learning outcome measures. Third, a visualization of the lab tests that warranted the main hypotheses was used to facilitate interpretation of the misconceptions. Fourth, decision rules were established to address misconceptions and improve user modeling processes.

**Extraction of patterns**

Table 2 shows the results obtained by the algorithm with the different parameter values, including the total number of rules obtained, and the values of the quality measures. For the purposes of this analysis, a low number of rules with few attributes is preferred in order to ease the interpretability of the results. Furthermore, the rules are expected to be indicative of common misconceptions when diagnosing diseases. Therefore, the desired result consists of rules that are both precise and accurate in recognizing incorrect lines of diagnostic reasoning on the basis of non-pertinent lab tests ordered by the learners.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect diagnosis for the Cynthia case (i.e., exhibiting signs of Pheochromocytoma)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Blood Glucose Level Result: normal=true</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>6.1%</td>
<td>83.3%</td>
<td>43.9%</td>
</tr>
<tr>
<td>Continual ECG Monitoring Result: Normal=true</td>
<td>9</td>
<td>2</td>
<td>11</td>
<td>11.2%</td>
<td>81.8%</td>
<td>46.9%</td>
</tr>
<tr>
<td>Fasting Blood Glucose Level Result: normal=true</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>7.1%</td>
<td>71.4%</td>
<td>42.9%</td>
</tr>
<tr>
<td>ECG with exercise Result: Normal=true</td>
<td>7</td>
<td>3</td>
<td>10</td>
<td>10.2%</td>
<td>70.0%</td>
<td>43.9%</td>
</tr>
</tbody>
</table>

*Note. 1 = Positive instances; 2 = Negative instances; 3 = Size; 4 = Coverage; 5 = Precision; 6 = Accuracy.*

It is apparent from this table that learners exhibit very few misconceptions while solving the case. A total of 4 rules were extracted to achieve the minimal coverage criteria of 5% of lines of diagnostic reasoning. The range of lines of diagnostic reasoning that were classified as characteristic of a common misconception varied from 6 (6.1%) to a total of 11 (11.2%) lines of diagnostic reasoning. It is noteworthy to mention that rules were only generated in relation to lines of diagnostic reasoning associated to the Cynthia case, which is recognized as the most difficult to solve as Pheochromocytoma is a rare disease. Furthermore, all lab results that are featured as antecedents within the rules are non-pertinent, meaning that the patient vital signs were found to be normal. In sum, misconceptions were only identified in solving more complex cases as opposed to simple ones, as learners attempted but failed to confirm their incorrect diagnoses.
The subgroup discovery rules reveal interesting information about learner misconceptions that can be useful in revising the learner model implemented in BioWorld. As an example, consider the first rule in Table 2. This rule shows that learners who ordered a test to measure Random Blood Glucose levels obtained a value that falls within the normal population range. On the one hand, there are a total of 5 lines of diagnostic reasoning that were characterized by this behavior and that led to an incorrect diagnosis. On the other hand, there was only 1 line of diagnostic reasoning where the learner engaged in this behavior, but still demonstrated a correct understanding of the underlying disease. As such, this particular rule is quite precise, but fairly accurate in characterizing all the lines of diagnostic reasoning included in the dataset (i.e., Precision = 83.3%; Accuracy = 43.9%).

From a medical point of view, higher than normal blood glucose levels may be attributed to rare tumors such as Pheochromocytoma, but also to other causes such as an overactive thyroid gland, pancreatic cancer, pancreatitis, and so on. The most relevant test, however, was to assess the levels of catecholamines through a urine test, as this particular type of tumor usually grows on the adrenal glands. This rule provides new information on how to improve the novice-expert overlay model in BioWorld. It suggests that learners who order this particular test need a decision support system to support them in performing a differential diagnosis. This involves BioWorld supporting the learner in (a) listing the relevant symptoms, (b) the possible diseases that explain the symptom, (c) ruling out possible causes, beginning by removing diagnoses from the list given the non-pertinent lab test result. In this particular case, a learner should rank the likelihood of arrhythmia as less likely given that high blood sugar levels influence heart rhythm. Therefore, arrhythmia would likely be the cause of the symptoms if the lab test showed elevated levels of blood glucose, but this was not the case. The novice-expert overlay model should be programmed to assess these behaviors if the learner orders this particular lab test.

Validation of patterns

Table 3 shows the performance data associated with the outcomes of the learners’ efforts to solve the problems in BioWorld. The performance metrics associated with their final solution includes the count and percentage of match with the expert solution, the count of lab test, the level of confidence towards the final solution, as well as the count of correct solutions, incorrect ones, and the ratio of these values. The metrics related to individual lines of diagnostic reasoning during problem-solving consist of the count of correct and incorrect lines of diagnostic reasoning as well as the ratio of these values. Given that the rules generated by the subgroup discovery algorithm are indicative of misconceptions, it is expected that these are associated with lower levels of efficacy, efficiency, and confidence in solving medical problems. In contrast, lines of diagnostic reasoning that are classified by the set of rules as demonstrating no misconceptions should be associated with higher levels of efficacy, efficiency, and confidence.

<table>
<thead>
<tr>
<th>Subgroup</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>
</tr>
</thead>
<tbody>
<tr>
<td>Cynthia case (i.e., exhibiting signs of Pheochromocytoma)</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>Misconceptions</td>
<td>18</td>
<td>49%</td>
<td>3.22</td>
<td>14.89</td>
<td>73%</td>
<td>6</td>
<td>12</td>
<td>0.50</td>
<td>7</td>
<td>11</td>
<td>0.64</td>
</tr>
<tr>
<td>No Misconceptions</td>
<td>80</td>
<td>55%</td>
<td>3.70</td>
<td>13.65</td>
<td>73%</td>
<td>33</td>
<td>47</td>
<td>0.70</td>
<td>50</td>
<td>30</td>
<td>1.67</td>
</tr>
<tr>
<td>Amy case (i.e., exhibiting signs of Diabetes Mellitus (type I))</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>No Misconceptions</td>
<td>105</td>
<td>43%</td>
<td>6.84</td>
<td>10.97</td>
<td>78%</td>
<td>59</td>
<td>46</td>
<td>1.28</td>
<td>79</td>
<td>26</td>
<td>3.04</td>
</tr>
<tr>
<td>Susan Taylor case (i.e., exhibiting signs of Hyperthyroidism (Grave’s disease))</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>No Misconceptions</td>
<td>101</td>
<td>71%</td>
<td>6.22</td>
<td>8.92</td>
<td>86%</td>
<td>58</td>
<td>43</td>
<td>1.35</td>
<td>85</td>
<td>16</td>
<td>5.31</td>
</tr>
</tbody>
</table>

Note. 1 = N; 2 = Average percentage of matches with expert solution; 3 = Average count of matches with expert solution; 4 = Average count of lab tests; 5 = Average level of confidence in case solution; 6 = Count of correct lines of diagnostic reasoning; 7 = Count of incorrect lines of diagnostic reasoning; 8 = Ratio of line correctness; 9 = Count of correct case solutions; 10 = Count of incorrect case solutions; 11 = Ratio of case correctness.

The differences between the subgroups that exhibit either a misconception or lack thereof are highlighted in Table 3 with respect to both the complex and simple cases solved in BioWorld. From this data, it is apparent that the subgroup that exhibited misconceptions when solving the most complex case performed less well than those solving simpler cases. The ratios of correct to incorrect lines of diagnostic reasoning are at least two times higher when solving simple cases. This value doubles to four times higher when appraising the ratios of the correctness of the final case solution for the simple cases. In general, learners who solve simple cases are more confident, order less lab
tests, and have higher counts of matches with the expert solution compared to the subgroup that exhibited misconceptions when solving the complex case.

Although this trend was evident when comparing the subgroup with misconceptions and no misconceptions in relation to solving a complex case, the trend clearly decreases from the levels of performance obtained by the subgroup that solved simple cases. The difference in performance was most notable in ratios correct to incorrect case solutions, where the subgroup with no misconceptions was at least two times more likely to obtain the correct solution. However, the difference between both subgroups is less pronounced in relation to the other performance metrics. There were only slight differences in efficacy, efficiency, and confidence between the subgroups that exhibited misconceptions as opposed to the one that did not while solving the complex case.

**Interpretation of patterns**

Figure 2 elaborates further on the nature of the misconceptions exhibited in the lines of diagnostic reasoning that were recognized by the subgroup discovery algorithm. In order to further characterize these misconceptions, we examined the main hypothesis selected at the end of each line of diagnostic reasoning that exhibited a misconception. Furthermore, we examined whether the evidence was linked with a particular diagnosis after solving the case, and how the evidence was categorized and prioritized by the learner before BioWorld delivered feedback. The thickness of the lines reflects the frequency count of learners who fell in each category, with thicker lines indicating higher frequencies.

*Figure 2. Interpretation of the patterns extracted by the subgroup discovery algorithm*

Three broad types of misconceptions emerged from this analysis. When solving a case exhibiting symptoms of Pheochromocytoma, the lines of diagnostic reasoning where a misconception was evident led to incorrect diagnoses of Panic Attack, Hyperthyroid, and Arrhythmia. However, six lines of diagnostic reasoning from a total of 23 within the subgroup that exhibited misconceptions were later followed by the correct diagnosis of Pheochromocytoma. Furthermore, the subgroup rarely engaged in behaviors that were indicative of using the non-pertinent lab test to
warrant erroneous diagnoses, as the frequencies of linking, categorizing, and prioritizing these evidences were relatively low. This finding suggests that although these misconceptions are commonly encountered while solving the complex case, they are not necessarily constant or immutable. Lines of diagnostic reasoning unfold as a dynamic process as the learners’ confidence in multiple diagnoses shift on the basis of available symptoms, lab tests, and library information. Given the fact that misconceptions are characterized as a transient condition, their early detection by BioWorld is the first step to delivering remedial instruction that will support learners in becoming more competent.

Conclusion

Learner modeling in computer based learning environments has become a central challenge for instructional technology designers and researchers. Learners encounter obstacles and hold misconceptions that can hinder learning. Therefore, it is imperative to support learners in their individual learning trajectories by embedding scaffolding and feedback through computer based cognitive tools. A chief obstacle to the design of adaptive learning environments is to first identify individual differences and to respond to misconceptions. The redesign of functionalities and response mechanisms in learning environments will substantially contribute to this objective.

The purpose of this study was to develop an understanding of misconceptions in medical problem solving by exploring pertinent and non-pertinent laboratory tests in BioWorld, with the goal of adapting instruction and providing requisite scaffolding. To address this purpose we used subgroup discovery to generate rules, which provided evidence about the relationship between misconceptions in clinical reasoning and laboratory tests ordered in BioWorld. The results suggested that subgroup discovery algorithms are particularly well suited towards ascertaining the difficulties that characterize problem-solving in ill-structured domains such as medical problem solving. In particular, the analyses revealed the specific antecedents of incorrect diagnoses, with regards to the laboratory tests ordered. Performance metrics were associated with correct performance as well as specific misconceptions for each of the cases. The results from this study have implications for evaluating and deriving empirically-based design guidelines with regards to tailoring instruction and scaffolding to individual learners. For example, these findings will contribute towards improving the metacognitive support provided to help learners become aware of misconceptions in solving medical problems.

It is worth considering some limitations of this investigation. First, there are many factors that can contribute to the decision of ordering certain lab-tests, of which we only considered a few. In future studies, we will systematically analyze learner rationales for ordering certain lab-tests by examining think out loud protocols. Second, we constrained our analysis to a small number of rules with few attributes to ease the interpretability of the results. We will conduct an expanded analysis to generate a complete set of rules with all the attributes to establish a holistic picture of misconceptions during medical problem solving. Third, our analyses were limited to three medical problems, which challenges the generalizability of our findings. Extracting patterns from a larger set of clinical cases will lead to a better understanding of misconceptions and more precise generation of rules across diverse medical problems.

The use of subgroup discovery is a promising method to reveal and understand the relationship between misconceptions in clinical reasoning and types of laboratory tests ordered. There is still much to be gained from examining the misconceptions that learners encounter and the ways to help support learners overcome these difficulties. Future work will consist of investigating alternative approaches for delineating the relationship between obstacles in clinical reasoning and lab-tests ordered. More importantly, we hope to incorporate the lessons gleaned from this study to make the novice-expert overlay model in BioWorld more robust to provide adaptive scaffolding for learners engaging in clinical reasoning. This work is an important step towards redesigning computer based learning environments to support learners on their individual learning trajectories.

Acknowledgements

We gratefully acknowledge the contributions of Laura Naismith and Maedeh Kazemitabar in collecting data for this project. This research is funded by the Social Sciences and Humanities Research Council of Canada.
References


ABSTRACT

Flow is the affective state in which a learner is so engaged and involved in an activity that nothing else seems to matter. In this sense, to help students in the skill development and knowledge acquisition (referred to as learners’ growth process) under optimal conditions, the instructional designers should create learning scenarios that favor the learner’s flow. One-way to obtain these scenarios is to incorporate the conditions postulated by the Flow Theory in the context of the instructional design process. However, little attention has been drawn to this integration, and how to apply the Flow Theory based on learning theories that provide theoretical justifications during the instructional design process. Thus, in this paper, we propose a framework to integrate the learner’s growth process and the Flow Theory. It provides adequate support for the instructional design of learning scenarios that lead and maintain learners in flow state. We demonstrate the usefulness of this framework by presenting an application that helps designers to search and select learning objects that have the potential to maintain the learner’s flow in a learning scenario.

Introduction

In the learning process, the learner’s affective state plays an essential role that influences several mechanisms of rational thinking and learning (D’Mello, 2012; Picard, 1997; Duque Reis et al., 2015). Learners citing negative affective states (e.g., boredom) during learning activities are, in general, significantly more likely to show inadequate learning outcomes, because they often are not motivated and are not engaged in learning (Craig, Graesser, Sullins, & Gholson, 2004; Shernoff, Vandell, & Bolt, 2008). To motivate a student so that he/she performs learning activities with complete immersion, it is necessary that his/her affective state provide an optimal experience. This affective state is denominated flow, and it is a mental state of operation characterized by a feeling of energized focus, full involvement, and success in the task being performed (Csikszentmihalyi, 2013). One condition for attaining and maintaining the flow state is a good balance between the perceived challenges of a task and the participant’s own perceived abilities to solve it (Csikszentmihalyi, 2014). A task that is too challenging or one that is not challenging enough may lead to frustration or boredom, hence demotivating learners.

Thus, to define a learning scenario with specific learning objects that favors and maintains the learner’s flow during educational activities, instructional designers need to have some understanding about the influence of these activities/objects in the affective state of a learner. For example, one factor that affects the perceived challenge of a given activity is the difficulty levels of learning objects. In this case, if the difficulty level of a learning object (or sequence of learning objects) is not adequately connected with the learning goals and the current knowledge and skills of a student, the learning scenario will be perceived as too hard and frustrating, or as too easy and boring.

In a collaborative learning scenario, the challenge of designing adequate activities and selecting learning objects is even harder. If the instructional designer selects problems and learning objects that are too difficult (or too easy) for students, it will hinder students’ interactions, demotivate students, and lead students to not want to work in groups over time (Challco, Moreira, Mizoguchi, & Isotani, 2014; Isotani, Inaba, Ikeda, & Mizoguchi, 2009). For instance, consider a scenario where a student (the tutor) interacts with another student (the tutee) to solve a given problem (i.e., a selected learning object). In this situation, the tutor will learn by using his knowledge/skills to demonstrate how to solve a problem and the tutee will learn by following the tutor’s guidance. If the problem is too hard or the sequence of activities is not created to help students to collaborate, the tutor will not have the sufficient skill level or...
knowledge to solve and guide the tutee in the resolution of the problem. As a result, the learning scenario will cause emotional distress in both tutor and tutee, and the desired learning outcomes will not be achieved.

To support the design of better learning scenarios that are pedagogically sound and can keep learners in a flow state, it is essential during the instructional design process to take into account the level of difficulty of learning objects and to link learning objects with theories that describe learners’ growth. Unfortunately, this task requires specialized knowledge about instructional/learning theories, Flow Theory, and Affect Theory, and the skills to apply this knowledge in an integrated manner in order to select adequate learning objects and design effective learning scenarios that match students’ abilities.

To support the design of authoring tools that help instructional designers with the proper selection of levels of difficulty that keep the learners in flow, in this paper we propose a framework to integrate the learner’s growth process and Flow Theory through a new theory-based model, named GMIF: Learner’s Growth Model Improved by Flow Theory. This model explains and describes the necessary conditions under which learners are able to learn more effectively based on learning theories, while keeping the ability-challenge balance of tasks defined in the Flow Theory. In particular, the GMIF has been used to create algorithms that help to automatize the selection of proper learning objects for specific learning situations.

This paper is organized as follows: The next section provides details about the Learner’s Growth Model (LGM) and Flow Theory. Then we present related works on the application of Flow Theory in educational settings. Following that, we present the GMIF, offering a detailed description of our framework that integrates the LGM and Flow Theory. Next, to demonstrate the usefulness of our framework, we illustrate how the GMIF can be used to design learning activities and select learning objects. Finally, we present the conclusion of our work and recommendations for future research.

### Theoretical background

#### Learner’s Growth Model

Based on learning theories, the Learners Growth Model (LGM) is a graph that represents the learning process of a student as stages of skill development and knowledge acquisition in a simplified way (Inaba, Ikeda, & Mizoguchi, 2003; Isotani & Mizoguchi, 2006). The learner’s growth process is represented as paths on the graph that allow for the representation of the relationships between learning strategies and their educational benefits.

As shown in Figure 1, the LGM has twenty states that are the result of the number of stages related to skill development multiplied by the number of stages related to knowledge acquisition. In the graph, the stages of skill development (nothing, rough cognitive, explanatory cognitive, associative, and autonomous) are represented in the lower-left triangle, while the stages of knowledge acquisition (nothing, accretion, tuning, and restructuring) are represented in the upper-right triangle. In skill development, the cognitive stage (rough and explanatory) involves an initial encoding of a target skill that allows the learner to present the desired behavior or, at least, some rough approximation thereof; the associative stage is the improvement of the desired skill through practice; and the autonomous stage involves gradual continued improvement in the performance of the skill (Anderson, 1982). During knowledge acquisition, the accretion stage incorporates the addition and interpretation of new information in terms of pre-existent knowledge; the tuning stage involves coming to understand the knowledge through its application in a specific situation; and the restructuring stage comprises a process in which the relationship of the acquired knowledge is considered and the existent knowledge structure is rebuilt (Rumelhart & Norman, 1976).

The arrows in the LGM shown in Figure 1 represent possible transitions between stages, and the form \( s(x,y) \) on the top of each vertex is the simplified form of representing a stage, where \( x \) represents the current stage of skill development and \( y \) represents the current stage of knowledge acquisition. For instance, the transition \( s(0,0) \rightarrow s(2,0) \) means the possible transition from stage \( s(0,0) \) where a learner does not have any knowledge or skills to the associative stage \( s(2,0) \) of skill development.
One of the most interesting uses of this model is the representation of learning strategies based on the benefits that different learning theories offer. Figure 2 shows two examples of this representation based on Cognitive Apprenticeship Theory (Collins, 1991), in which each black arrow implies the application of a learning strategy that facilitates the learner's growth process. Figure 2 (a) shows the application of the apprenticeship learning strategy, where a student grows his/her cognitive skill from nothing $s(0,y)$ into the associative stage $s(3,y)$ through the rough cognitive stage $s(1,y)$ and the explanatory cognitive stage $s(2,y)$. Figure 2 (b) shows the application teaching learning strategy, where a student grows his/her cognitive skill from the associative stage $s(3,y)$ into the autonomous stage $s(4,y)$. With the use of the LGM, any learning strategy or educational best practice can be explicitly described as a path on the graph, facilitating the understanding, visualization and utilization of the model (Isotani, Mizoguchi, Inaba, & Ikeda, 2010).

Flow Theory

Csikszentmihalyi’s Flow Theory constitutes an important theory regarding the study of affective states during activities that require active work from participants, such as discussions, exercises, and group activities (Csikszentmihalyi, 2014; Snyder, Lopez, & Pedrotti, 2010). This theory has been applied in several fields, including
game design, commerce, and education. The key concept of this theory is the definition and description of “The Zone” (i.e., flow state) as a situation where a person is so engaged and focused on a particular task that he/she is completely immersed in it (Csikszentmihalyi, 2013). According to Flow Theory, the following conditions must be satisfied to achieve flow:

- Clear goals in which the expectations and rules are clearly discernable
- Direct and immediate feedback in which the successes and failures of the task are apparent, so that behavior can be adjusted as needed
- Good balance between ability level and challenge

One of the conditions given above is that the flow state only occurs if there is a good balance between the perceived challenges of the task at hand and the learner’s own perceived ability to solve it. This means that the definition of an appropriate challenge (i.e., level of difficulty) is fundamental to the design of situations that promote a flow state (Linehan, Bellord, Kirman, Morford, & Roche, 2014). Thus, Csikszentmihalyi proposes the three-channel model of Flow Theory (Csikszentmihalyi, 2013) shown in Figure 3, in which both anxiety and boredom drive persons to frustration. When a task is too difficult to be solved, it causes anxiety because it is perceived as too challenging or because the person’s skill level is not sufficient to solve the task. In the same way, when a task is too easy it causes boredom because it is not challenging enough, or because the person’s skill level is too high for the task.

![Figure 3. Affective state in terms of ability level and challenge level, according to the Flow Theory model](image)

In the context of education, learners in the flow state frequently experience positive affect and better scores/performances compared with other learners who are in a similar situation but not in flow (Baydas et al., 2015; Ibáñez, Di Serio, Villarán, & Delgado Kloos, 2014; Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2003). For example, several empirical studies conducted by D’Mello and Graesser, using intelligent tutoring systems (e.g., AutoTutor) have shown strong positive correlations between learning gains, confusion, and flow (D’Mello & Graesser, 2012). Another example is the study of Choi et al., (2007), where participants used a web-based e-learning system in a program on Enterprise Resource Planning. The data of this study revealed that flow experiences were directly linked with learning outcomes and learners’ attitude towards e-learning.

**Related works**

The Flow Theory and the three-channel model proposed in Csikszentmihalyi (2013) have been frequently used for building instruments and tools for detection of the flow state (Kort, Reilly, & Picard, 2001; Pearce, Ainley, & Howard, 2005; Vanlehn, 2006; Shernoff et al., 2008; Esteban-Millat, Martínez-López, Huertas-García, Meseguer, & Rodríguez-Ardura, 2014). More recently, studies have attempted to analyze and model students’ flow experiences in order to (a) evaluate students’ interactions with learning objects; (b) personalize educational activities (e.g., lessons); and (c) develop better content. In this arena, the study of Del Blanco et al. (2012) proposes a framework to support the integration of games in learning activities. To do so, authors identified key aspects about the mechanisms that facilitate the use of pedagogical approaches with games to keep students in flow state. Then they proposed a
workflow for integrating games into learning designs. As a result, it is possible to create guidelines to help educators use (and reuse) sequences of games in the learning process. Although this work provides some initial support for creating better learning experiences, if the games themselves do not have the qualities and attributes necessary to maintain student engagement, flow experiences will not occur. Considering this problem, Kiili et al. (2014) presented a framework for analyzing and designing educational games based on Flow Theory. This framework describes several dimensions of flow experience as well as meaning factors that affect the design of game-based learning activities. Still, it is limited to specific parts of the game design process.

The related work and frameworks presented in the previous paragraph are important for guiding educators and game designers to create better learning situations. Nevertheless, they were not created to automate the process of learning design and do not have the necessary formalization to be implemented and included as a feature in a learning design authoring tool. This means that if an instructional designer wants to support flow, he/she will need to do so manually, without any computational support. Such a manual approach is infeasible to be carried out when there is a need to plan personalized sequences of activities for a class of students with different needs, using a database with multiple learning objects (e.g., games, texts, videos, images) and taking into consideration several pedagogical approaches to support flow experiences. Toward the automation of detecting and using flow state to create better learning experiences, Lee and colleagues provide adaptive learning contents by selecting appropriate problems based on Flow Theory (Lee, Jheng, & Hsiao, 2014). They propose an automatic flow detector where the three-channel model of flow is built based on features related to affective dimensions (i.e., valence and arousal) and interactions (i.e., mouse click duration, keystroke duration) with learning software.

As we can observe, despite the broad use of Flow Theory in educational contexts, to the best of our knowledge, there is not a computational model/framework based on Flow Theory that provides support for instructional designers to create learning scenarios that maintain the flow experience while offering theoretical justifications regarding the learner’s growth. In particular, there is no computational help to define appropriate challenges for learners and to select adequate learning objects as needed to promote a flow state. Thus, our approach aims to fill this gap by proposing a computational framework that connects Flow Theory and the learner’s growth process. It allows for identifying a proper balance between ability and challenge according to the learner’s state of knowledge acquisition and skill development. It can also be used to create algorithms to equip authoring tools with intelligent features to automatize the design of learning scenarios that increase the chances of students reaching and staying in flow state.

**Framework to integrate the learner’s growth process with Flow Theory**

The tasks that will be performed by a student in a pedagogical scenario are learning activities. The ability-challenge balance defined in Flow Theory should be determined by the current stage of the learner’s growth process and the difficulty level to maintain the learner’s flow. Therefore, our framework integrates the learner’s growth process and Flow Theory by labeling each arrow $s(x,y) \rightarrow s(x,y)$ of the LGM with the form $(z_{min}, z_{max})$, indicating the minimum level ($z_{min}$) and the maximum level ($z_{max}$) of difficulty that are necessary to maintain the learner’s flow. This LGM with labels is called the GMIF: Learner’s Growth Model Improved by Flow Theory.

In this section, we first show how a scale of five levels of difficulty is obtained with the GMIF. Next, we present our framework as a pseudo-algorithm that can be used to create a GMIF of any scale (n-scale GMIFs).

**Five-point scale GMIF**

As we see in the three-channel model of Flow Theory, the perceived ability and challenge are indicators used to identify a person’s affective states. These two indicators are represented in the model of Flow Theory as an axis without scales and labels to depict situations where a learner can be anxious, bored or in a flow state. Thus, first we divide the axis into five parts to obtain a five-point scale. Next, the transitions in the skill development or knowledge acquisitions defined by the LGM model are set for the ability axis as the scales of space using a uniform distribution.

Figure 4 shows the five-point scale of skill-difficulty that is defined by employing the five levels of difficulty [0: very-easy, 1: easy, 2: medium, 3: difficult, 4: very-difficult], and the transitions in skill development [$s(0, y) \rightarrow s(1, y)$: from nothing to the rough-cognitive stage; $s(1, y) \rightarrow s(2, y)$: from the rough-cognitive to the explanatory-cognitive
stage; \( s(2,y) \rightarrow s(3,y) \): from the explanatory-cognitive to the associative stage; \( s(3,y) \rightarrow s(4,y) \): from the associative to the autonomous stage. According to this five-point scale, the label sequence for maintaining the learner’s flow is \( s_1 = \{ [0;0], [1;2], [2;3], [4;4] \} \), in which each position represents a transition in skill development. The first element means that, during the transition \( s(0,y) \rightarrow s(1,y) \), the proper level of difficulty to maintain the learner’s flow is 0:very-easy. The second element means that, during the transition \( s(1,y) \rightarrow s(2,y) \), the proper level of difficulty to maintain the learner’s flow is 1:easy or 2:medium. The third element means that, during the transition \( s(2,y) \rightarrow s(3,y) \), the proper level of difficulty to maintain the learner’s flow is 2:medium or 3:difficult. Finally, the fourth element means that, during the transition \( s(3,y) \rightarrow s(4,y) \), the proper level of difficulty is 4:very-difficult.

By employing the transitions for knowledge acquisition (\( s(x,0) \rightarrow s(x,1) \): from nothing to the accretion stage; \( s(x,1) \rightarrow s(x,2) \): from the accretion to the tuning stage; \( s(x,2) \rightarrow s(x,3) \): from the tuning to the restructuring stage), we obtained the five-point scale of knowledge-difficulty as shown in Figure 5. In this space, the label sequence \( s_2 = \{ [0;0], [1;3], [4;4] \} \) means that, during the transition \( s(x,0) \rightarrow s(x,1) \), the level of difficulty should be 0:easy to maintain the learner’s flow. The second element in the sequence means that, during the transition \( s(x,1) \rightarrow s(x,2) \), the proper level of difficulty to maintain the learner’s flow is 1:easy, 2:medium, or 3:difficult. Finally, the proper level of difficulty during the transition \( s(x,2) \rightarrow s(x,3) \) is 4:very-difficult.

To obtain the five-point scale GMIF, we need to understand the relationship between the student’s growth state (i.e., perceived ability) and the difficulty level of learning materials (i.e., challenge). With this knowledge it is possible to design learning scenarios that (i) favor the maintenance of a flow state for students; and (ii) help them to achieve desired educational goals (i.e., acquisition of knowledge or development of skills). To accomplish our objective we used the analysis presented in Figures 4 and 5, which enables us to understand when students are in flow state (by making the correlation between knowledge and skills with a five-point difficulty level scale), to adequately label each transition (i.e., \( s(x,y) \rightarrow s(x',y') \)) between states in the LGM with a tuple \([z_{\text{min}}, z_{\text{max}}]\), where \( z_{\text{min}} \) refers to the minimum difficulty level necessary for a learning material (or sequence of materials) to be considered interesting and not too easy for a student, and \( z_{\text{max}} \) refers to the maximum difficulty level possible for the same material to be considered challenging but not too difficult. According to previous findings, such an explicit formalization is one of the key features needed to promote more effective and robust learning in e-learning environments (Esteban-Millat et al., 2014; Fulmer, D’Mello, Strain, & Graesser, 2015; Linehan et al., 2014). To define the values of each label \([z_{\text{min}}, z_{\text{max}}]\) for each transition \( s(x,y) \rightarrow s(x',y') \), we used the sequences \( s_1 \) and \( s_2 \) that define the proper levels of difficulty for each transition related to skill development and knowledge acquisition, respectively. Thus, when a transition is related to skill development, we used sequence \( s_1 \) and Figure 4. For example, to develop skills from the explanatory-cognitive stage to the associative stage, we label the transition \( s(2,y) \rightarrow s(3,y) \) in the LGM graph as \([2;3]\) by looking at
where this transition is located in the flow area presented in Figure 4. In this particular situation, [2;3] means that to maintain the learner’s flow, learning materials should be selected to challenge students by offering moderate levels of complexity to solve a task. Following the same procedure, the transitions related to knowledge acquisition use sequence S2 and Figure 5 to define the values of the labels. Figure 6, then, shows the five-point scale GMIF that results from labeling the LGM with a scale of five levels of difficulty.

![Figure 5. The five-point scale of knowledge-difficulty levels](image)

![Figure 6. Five-point scale Learner’s Growth Model Improved by Flow Theory](image)
There are several theoretical and practical applications of the proposed framework. One of them is helping instructional designers and educators to adequately select learning strategies (Figure 2) and challenges to create sequences of activities that have the potential to support the learner’s growth (based on learning theories) and at the same time provide a learning environment suitable for getting students into a flow state. It is also possible to use this framework for selecting learning objects that offer the right amount of challenge to students. We will further discuss the benefits of this framework in the following sections.

It is also important to note that although in this work we use a five-point scale of difficulty level, there are other scales of difficulty that can be considered during the design of learning scenarios. Thus, in the next section we will present how our framework can be mapped and used with any desired scale of difficulty.

**Pseudo-algorithm for n-scale GMIFs**

Figure 7 shows the pseudo-algorithm of our framework that can be used to build a GMIF for any difficulty level scale (n-scale GMIF), where the expected difference for levels of difficulty in the flow area is passed as a second argument (delta that has the default value zero). In the pseudo-algorithm, the symbol GMIF( (x,y), (x’,y’) ) represents a label for the transition from s(x,y) to s’(x’,y’) of the LGM, and each label is represented as the form (z_{min}, z_{max}), which indicates the minimum (z_{min}) and maximum (z_{max}) levels of difficulty for maintaining the learner’s flow.

The flow region in the pseudo-algorithm for the two areas of skill-difficulty and knowledge-difficulty are obtained by the function get_flow_region (lines 2-3), where the first parameter is the number of transitions (for skill development or knowledge acquisition), the second parameter is the scale of space, and the third parameter is the expected difference for levels of difficulty.

Figure 7. A pseudo-algorithm to build the n-scale GMIF

In the pseudo-algorithm shown in Figure 7, because the transition of skill development includes a flexibility that allows for increasing the skill stage without following all the transitions between stages, transition s(x,y) to s(x’,y’) (lines 4-5) is labeled by all levels of difficulty that are defined in intermediate transitions. For example, it is possible to go from 0:nothing to 3:associative stage without moving through 1:rough-cognitive stage and 2:explanatory-cognitive stage; thus, the transition s(0,y) to s(3,y) is labeled with the union of levels of difficulty defined in the transitions s(0,y) to s(1,y), s(1,y) to s(2,y) and s(2,y) to s(3,y). In the case of knowledge acquisition, the transition of stages is completed step-by-step without skipping any of the stages; thus the transition s(x,y) to s(x,y’) (lines 6-7) is labeled by setting the corresponding levels of difficulty knowledge-difficulty.

Figure 8 details the pseudo-algorithm for the function get_flow_region. This function calculates the flow region for skill-difficulty and knowledge-difficulty, where the flow region is represented as an array of size m (number of transitions for skill development or for knowledge acquisition) in which each i-th element contains the levels of difficulty for the transition from the i-th stage to the next stage (i+1 stage). For an instance of five levels of difficulty and three transitions of knowledge acquisition (shown in Figure 5), the flow region as a result of the pseudo-algorithm is a sequence \( s = \{(0;0),(1;3),(4;4)\} \), where the first element (0;0) indicates the level of difficulty as 0:very-easy for the transition s(x,0) to s(x,1).

The pseudo-algorithm to obtain the flow region shown in Figure 8 can be summarized in a narrative form as follows: Calculates the number of levels that should be distributed for each transition of stage (lines 3-12). These values are calculated through a uniform distribution that tries to maintain the same number of levels in all stages. The stages located in the same distance of mean stage should have the same number of levels. For example, the distribution of
eight levels of difficulty in five transitions is defined as the array \( s = \{1,2,2,2,1\} \), where the second, third, and fourth transitions are set with two levels \( s(1) = s(2) = s(3) = 2 \), while the first and fifth transitions are set with one \( s(0) = s(4) = 1 \). The transition located in the third transition is set with two levels \( s(2) = 2 \). The steps that calculate these levels are as follows:

- The normalization for the number of levels of difficulty. This is done to avoid the non-uniform distribution that happens when this number is odd and it is greater than the number of transitions. For example, the distribution of nine levels among four transitions only can be done by setting one transition with three levels, and setting the rest of transitions with two levels. Therefore, the normalization of level of difficulty is done by reducing the number of levels by one (lines 2-5). In the previous example, the distribution of nine levels into four transitions can be defined as the array \( s = \{2,2,3,2\} \) before the normalization, while the distribution of these nine levels after the normalization is defined as the array \( s = \{2,2,2,2\} \).

- After the normalization, the minimum number of levels of difficulty for each stage is defined by the function \texttt{initialize} (line 7), which initializes an array of size \( m \) with the value \( \lfloor n/m \rfloor \). The remaining levels of difficulty (line 8) are distributed according to the position \( \texttt{inv} \_\texttt{sigma} \) (lines 11-12). The value of \( \texttt{inv} \_\texttt{sigma} \) is the result of dividing the number of free spaces after the distribution of the remaining levels of difficulty by two (line 12).

```plaintext
get_flow_region(m, n\_difficulty=5, delta=0)
// normalize the number of difficulty levels
n ← n\_difficulty
if (n\_difficulty > m) and (odd? n\_difficulty) then
n ← n-1
// sets number of difficulty levels for each transition
distr ← initialize_array(m,\lfloor s/m \rfloor)
rest ← s - m \lfloor s/m \rfloor
if (rest > 0) then
inv\_sigma ← (n - rest)/2
For i from 0 to rest - 1 do
distr[inv\_sigma+i] ← distr[inv\_sigma+i] + 1
// make labels for flow region
flow[0].min ← 0
flow[0].max ← distr[0]-1
For i from 1 to m-1 do
flow[i].min ← flow[i-1].max + 1
flow[i].max ← flow[i-1].max + distr[i]
if (n\_difficulty > m) and (odd? n\_difficulty) then
flow[i].max ← flow[i].max + 1
if (even? m) then
if (i = \lfloor m/2 \rfloor) then
flow[i].max ← flow[i].max + 1
if (i = \lfloor m/2 \rfloor) then
flow[i].min ← flow[i]-1
Foreach r in flow do
if (r.max = -1) then r.min ← -1
else
r.min ← r.min-delta
r.max ← r.max+delta
return flow
```

Figure 8. A pseudo-algorithm to obtain a flow region in \( m \) transitions with \( n \) difficulty levels

After determining the number of levels of difficulty that will be distributed for each transition (\( \texttt{distr} \)), the next step is the setting of labels for the flow region that has no expected difference in the levels of difficulty (lines 14-24). Thus, the process to define these labels consists of:

- Setting of flow region for the first transition through the definition of minimum level with the value zero (line 14) and the definition of maximum level with the number of level of difficulty decreased by one (line 15).

- Setting of flow region for the rest of the transitions (lines 16-24). The minimum level of difficulty is defined as the maximum level of the previous transition increased by one (line 17), and the maximum level of difficulty is defined as the maximum level of difficulty of the previous transition increased by the number of levels (line 18). For cases in which the normalization of levels has been done, the following two rules must be applied:
If the number of transitions is odd, then the maximum level of difficulty is increased by one in the mid-transition (lines 20-21). Thus, the flow region for nine levels of difficulty in five transitions is defined as the array $s = \{(0;0),(1;2),(3;5),(6;7),(8;8)\}$.

If the number of transitions is even, then there are two mid-transitions: the first mid-transition is located in the position $\lfloor m/2 \rfloor - 1$, and the second mid-transition is located in the position $\lfloor m/2 \rfloor$. Next, the maximum level of difficulty is increased by one in the first mid-transition (line 23). Finally, the minimum level of difficulty is decreased by one for the second mid-transition (line 24). Thus, the flow region for nine difficulty levels in four transitions is defined as the array $s = \{(0;1),(2;4),(4;6),(7;8)\}$.

Finally, the expected difference in level of difficulty, defined as the parameter delta, is used to decrease and increase the minimum and maximum levels of difficulty for each transition in the flow region (lines 25-29).

Benefits and application of GMIF in learning design

Several factors must be considered during the learning design process, such as learning goals, pedagogical preferences, intervention timing, type of feedback, students’ needs, available resources, and so on. The work of Koedinger et al. (2013) estimates that there is a poll of $3^{30}$ (205 trillion) instructional choices that could be considered when designing a learning activity. Unfortunately, most designers and educators do not have enough knowledge/skills to cope with this huge number of instructional choices and select those choices that are the best fit for a particular situation.

To help in this process, our framework offers an appropriate integration of instructional design with learning theories, models of learner’s growth, and theories related to people’s experiences and engagement, namely Flow Theory, in order to reduce the complexity of the learning design task. Specifically, this section presents two concrete scenarios where designers can use GMIF to create optimal learning scenarios.

Foster flow experiences in theory-based learning scenarios

To get students into the flow state and produce optimal learning experiences, one should initially consider:

- The learner’s initial state and learning objectives in terms of knowledge acquisition and skills development (Anderson, 1982; Rumelhart & Norman, 1976);
- The learning path based on theoretical justifications (Isotani et al., 2010; Romiszowski, 1984); and
- The definition of the challenge. Here it is necessary to select the difficulty of the activities to keep students motivated and engaged (Csikszentmihalyi, 2013; D’Mello & Graesser, 2012).

We have developed the GMIF to support each of these steps. In the first step, the GMIF provide a standard to describe and represent learning objectives as well as the learner’s state. Thus, the problem of sharing learning designs among people (and computers) is reduced. Accordingly, an instructional designer can indicate the initial state of the learner and select his/her learning objectives. Both correspond to a state in the GMIF. After that, the designer can check manually or automatically (using learning design authoring tools) which learning theory (or theories) provides an adequate sequence of learning activities (i.e., learning path) that supports a learner in achieving the desired goals. In this situation, the GMIF offers visual representation (as a path on the LGM graph) of the theories and how they support the learner’s growth. Finally, to create flow experiences, the designer needs to define the difficulty of the challenge that the learner will work on. In this regard, the GMIF will indicate the level(s) of difficulty that should be considered when creating tasks to alter the state of the learner while keeping him/her motivated and engaged.

Develop flow-aware authoring tools to automate search and selection of learning objects

With the GMIF detailed in this paper, we can develop different functions for authoring tools for learning scenarios. A useful function is searching for proper learning objects that will favor and maintain the learner’s flow in a learning scenario. Thus, an instructional designer should first set the initial and goal stages of a student in a learning scenario using the graphical representation of the GMIF. Next, each label for a difficulty level in the transition from the initial
stage to the goal stage can be used as a constraint search. Finally, this constraint can be employed through a search process of different learning object repositories.

Ideally, the repositories should be able to search learning objects by the level of difficulty related to skill development or knowledge acquisition separately. However, the current repositories, following the standard for Learning Objects Metadata (LOM) (Thropp & McKell, 2001), only provide the field “educational.difficulty” in the metadata to define the levels of difficulty for learning objects. Thus, in this example application, we assume that this field (educational.difficulty) is related to both levels of difficulty (skill development and knowledge acquisition).

For the instance shown in Figure 9, where the application employs a five-point scale GMIF, if the instructional designer set the initial stage of a student as \( s(0,0) \)—nothing for skill and nothing for knowledge—and the goal stage as \( s(3,0) \)—associative stage for skill development—the proper levels of difficulty that will favor and maintain the learner’s flow are first, level 0: very easy for the transition \( s(0,0) \rightarrow s(1,0) \). Next, for the transition \( s(1,0) \rightarrow s(2,0) \), the proper levels of difficulty for learning objects that will maintain the learner’s flow are 1: easy, or 2: medium. Finally, for the transition \( s(2,0) \rightarrow s(3,0) \), the proper levels of difficulty for learning objects that will maintain the learner’s flow are 2: medium, or 3: difficult.

The solution shown in Figure 9 was obtained for a learning scenario based on the Cognitive Apprenticeship Theory (Collins, 1991), where a student without knowledge or cognitive skill should follow the strategy of learning by apprenticeship to reach the associative stage of cognitive skills. In this theory, a student using the strategy learning by apprenticeship will slowly improve his/her skills from nothing \( s(0,y) \) to associative stage \( s(3,y) \), moving through the rough cognitive stage \( s(1,y) \) and the explanatory cognitive stage \( s(2,y) \). In other words, this learning strategy is defined by the sequence \( s(0,y) \rightarrow s(1,y) \rightarrow s(2,y) \rightarrow s(3,y) \); more about the representation of this theory is presented in the works of Isotani and colleagues (Isotani et al., 2010; Isotani & Mizoguchi, 2006). Thus, when an instructional designer selects a transition in the graphical representation of the GMIF, the application automatically show a list of adequate learning objects (shown in Figure 9, left-bottom) that can be selected to maintain the learner’s flow during the selected transition. In Figure 9, a list of adequate learning objects that can be selected to maintain the learner’s flow during the transition from the rough cognitive stage \( s(1,0) \) to the explanatory cognitive stage \( s(2,0) \) is shown at the bottom of the graphical representation of the GMIF. The left portion of Figure 9 also shows the five-point scale of skill-difficulty that was used for the generation of the GMIF, where the label for the transition \( s(1,0) \rightarrow s(2,0) \) is emphasized by a double blue line. Furthermore, each dashed arrow from the five-point scale to the graphical representation of GMIF points to the corresponding label of levels of difficulty.
Conclusions and future research

Employing Flow Theory in the instructional design of learning scenarios is a way to favor the learner’s flow state in those scenarios. To incorporate Flow Theory in the design instructional process, a theory-based model that integrates the learner’s growth process and Flow Theory was proposed in this paper. This new model, called GMIF (Learners Growth Model Improved by Flow Theory) was developed by integrating the LGM (Learner’s Growth Model) and Flow Theory.

To demonstrate the utility of the current version of our model, we illustrated in this paper how an application can provide support to help instructional designers in the proper selection of learning objects that help to keep the learner in flow. We believe that the results shown in this paper can be used in theory-aware authoring systems, such as CHOCOLATO (Isotani et al., 2013), to improve the development of collaborative learning scenarios that favor the learner’s flow. In addition, we believe that introducing cognitive skills and knowledge level complexity into the LOM of the repositories would facilitate the use of the GMIF and even the LGM. In future work, our model should include others conditions of Flow Theory to provide better learning scenarios. Furthermore, we will develop empirical studies to validate our proposed model.

References


Identifying Effective Design Features of Technology-Infused Inquiry Learning Modules: A Two-Year Study of Students’ Inquiry Abilities

Ying-Shao Hsu1,2*, Su-Chi Fang1, Wen-Xin Zhang1, Hsin-Kai Wu1, Pai-Hsing Wu1 and Fu-Kwun Hwang3

1Graduate Institute of Science Education, National Taiwan Normal University, Taiwan // 2Department of Earth Sciences, National Taiwan Normal University, Taiwan // 3Department of Physics, National Taiwan Normal University, Taiwan // yshsu@ntnu.edu.tw // suchhii@msn.com // haha1788@gmail.com // hkwu@ntnu.edu.tw // brain1124@gmail.com // hwang@phy.ntnu.edu.tw

*Corresponding author

(Submitted January 26, 2015; Revised June 17, 2015; Accepted July 29, 2015)

ABSTRACT

The two-year study aimed to explore how students’ development of different inquiry abilities actually benefited from the design of technology-infused learning modules. Three learning modules on the topics of seasons, environmental issues and air pollution were developed to facilitate students’ inquiry abilities: questioning, planning, analyzing, and modeling. Two classes of seventh graders (13-14 years old) were randomly assigned to an inquiry group (n = 24) and a baseline group (n = 27). The two groups engaged in three technology-infused learning modules and conventional instruction respectively, and took two inquiry ability tests three times throughout two years. The test results showed that the inquiry group performed significantly better than the baseline group with respect to several inquiry sub-abilities. The instructional components of the tasks (i.e., design features) for which students exhibited significant improvements were further analyzed with respect to the task complexity, the function of the prompts, and the types of representations presented in the tasks. These design features identified provided useful insights into future design of technology-infused learning tasks for different inquiry abilities.

Keywords

Inquiry abilities, Instructional design, Technology-infused learning, Scientific inquiry, Learning module

Introduction

Scientific inquiry has become the focus of reforms in science education over the past decades. The use of computers in science learning has opened ample opportunities for inquiry. Computers offer tools that support collecting or analyzing data, provide simulations that make an extremely large/small-scale phenomenon available for investigation, and allow learners to test and describe their ideas through a modeling tool (de Jong, 2010; van Joolingen, de Jong, & Dimitrakopoulou, 2007). Relevant research has also shown that students’ science learning could benefit from their engagement in computer-based inquiry environments (de Jong, 2006a; Fretz et al., 2002; Quintana et al., 2004; van Joolingen, 1998).

Despite the many merits that computers can offer, researchers emphasized that “it is the tasks, activities and underpinning pedagogical strategies supported or facilitated by the technology rather than the technology itself that have an impact on learning” (Dalgaro & Lee, 2010, pp. 17-18). Diverse instructional designs that involve the use of computers were proposed for supporting students’ learning of scientific inquiry. For example, some researchers suggested designing the interface of the computer-based learning environment in a way that can help reduce the complexity of the inquiry learning tasks (Reiser, 2004). Others recommended incorporating interactive simulations into learning tasks, by which learners can be guided to test their ideas by manipulating simulated or real data. It is expected that interactive simulations would help learners make sense of the data and build arguments based on evidence (Hsu, 2008; McNeill, 2011).

These various instructional designs in computer-supported learning environments provide affordance to support learners’ inquiry learning; however, previous research has indicated that there is a gap between what curriculum designers intended to offer and what learners actually perceived and used (Wu, Wu, Kuo, & Hsu, 2015). The curriculum designers or teachers assume that the design features are explicit but learners might not notice and appreciate the features intended to present thus could not complete the task in a desirable way. In other words, it is not clear how these various instructional designs actually influence learners’ development of different inquiry
abilities, and in what circumstances the design features can be best applied (van Joolingen et al., 2007). Therefore, the study was intended to explore how instructional designs impacted on students’ learning of different scientific inquiry abilities. We developed three computer-based inquiry learning modules that provided tools to support students’ learning of several specific inquiry abilities. By means of a scientific inquiry test, we examined how these learning modules helped the development of different inquiry abilities. Then, through the lens of students’ learning outcomes, we identified design features of the inquiry tasks that might contribute to students’ progress of the inquiry abilities. The design features identified in the study provide useful guidance for future instructional designs in computer-based inquiry environments.

Design of computer-based inquiry learning modules

The study developed three inquiry learning modules on the topics of Seasons, Environmental issues, and Air Pollution. With reference to previous literature and considering the learning topics selected in the study, we adopted three overarching aspects to design the learning activities. The three overarching aspects are decomposing open-ended problems, providing prompts, and using interactive visualization tools.

Decomposing open-ended problems

One significant purpose of scientific inquiry is to make science meaningful for learners. To meet the purpose, educational researchers argue that learners need to be engaged in the practice of science in order to learn how scientists work in the discipline of science (Flick & Lederman, 2006). They also suggest that science learning and learners’ everyday world experience need to be connected (Millar & Osborne, 1998). A real-world problem not only provides an everyday-life-related context to involve learners in the process of investigation (Lehrer & Schauble, 2006) but also enables to organize a series of scientific practices such as scientific modeling, argumentation, and explanation in a coherent and meaningful way. We purposely selected three everyday-life-related topics for the learning modules: seasons, environmental issues and air pollution. For each learning module, we proposed an open-ended question to motivate students’ inquiry learning. The learning activities and tasks were then designed revolving around the specific question. That is, the learning activities and tasks were designed with the aim to answer the question being proposed.

Scientific investigations are complex processes. Therefore, for learners, learning science through inquiry is indeed demanding (de Jong, 2006b; van Joolingen et al., 2007). The challenges stem from cognitive complexity of the practices as well as unfamiliar use of language and new social interaction practice (Reiser, 2004). In order to reduce the complexity of inquiry tasks, researchers suggest decomposing open-ended problems and structuring component tasks in a logical way that can help guide learners to complete the learning process. For example, the Model-It software uses three functional modes: plan, build and test, to structure the task of modeling. The arrangement constrains the space of learning activities by breaking down the complex task into three parts, and helps learners construct a quality model step by step (Quintana et al., 2004). In this study, the questions posed in the modules were complicated in nature and involved multiple variables. Therefore, in order to decrease task complexity, we decomposed the multiple-variable problems into several activities that concentrated students’ attention on one single variable at a time.

Providing prompts

Understanding science as inquiry is challenging for middle school students as it involves an extensive range of cognitive and social skills that should be developed during this developmental stage (Bybee, 2006; Flick, 2006; McNeill, Lizotte, Krajcik, & Marx, 2006). Students do not naturally develop these competencies; they need support and guidance when learning and practicing scientific inquiry (Osborne, Erduran, & Simon, 2004). Researchers have explored various ways to scaffold students in inquiry learning, such as software tools and written prompts (e.g., Fretz et al., 2002; McNeill et al., 2006), and this work provided rich examples of promising design approaches and research. Based on the difficulties students faced during inquiry tasks, previous studies established frameworks for designing cognitive and meta-cognitive scaffolds to support inquiry learning (Quintana et al., 2004; Quintana, Zhang, & Krajcik, 2005). These frameworks provided rational and approaches for designing pedagogical support.
For example, students might not know how to collect data that are relevant to the research question. We can use prompts to guide students to think about the reasons for selecting certain variables for collecting data (Chang, Hsu, Wang, & Ho, 2015). In light of the frameworks, the study designed and embedded different prompts in the learning tasks, which helped students to clarify the task goal, draw their attention to the key elements to solve the problem, and support them to construct thorough scientific explanations.

Using interactive visualization tools

Research has shown that using visualization technologies in inquiry learning significantly enhance student science learning (de Jong & van Joolingen, 1998; Linn, Davis, & Bell, 2004; Quintana et al., 2004; White & Frederiksen, 1998). Visualization tools not only help learners “visualize” unobservable scientific phenomena such as molecular structures (Chang & Linn, 2013; Wu, Krajcik, & Soloway, 2001) but also empower learners to analyze and visualize data by transforming and communicating among multiple representations (Chang, 2013). In addition, interactive visualization tools can be used for supporting learners in constructing and revising models (Baek, Schwarz, Chen, Hokayem, & Zhan, 2011; Fretz et al., 2002) by which learners can manipulate variables and observe resulted patterns or causal relationships among relevant variables. In this study, we used three different visualizations tools: SeasonSim (Hsu, 2008), My World GIS, and Air-Pollution Modeling Tool (APoMT) software (Wu et al., 2015), to support students’ investigations in these three learning modules. In the next section, we will describe in more detail about how these visualization tools were used in each learning module.

Instructional design

In reference to Inquiry and the National Science Education Standards (INSES) (National Research Council, 2000), the study focused on four inquiry abilities: questioning, planning, analyzing, and modeling. In order to set and examine precise learning goals, we broke down each inquiry ability into several sub-abilities (see Table 1). These sub-abilities formed the key constructs of the development of learning modules and assessments.

<table>
<thead>
<tr>
<th>Inquiry abilities</th>
<th>Sub-abilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questioning</td>
<td>Q1. Posing a testable question</td>
</tr>
<tr>
<td></td>
<td>Q2. Recognizing key variables of a question</td>
</tr>
<tr>
<td>Planning</td>
<td>P1. Designing an experiment with relevant variables</td>
</tr>
<tr>
<td></td>
<td>P2. Controlling and manipulating variables effectively</td>
</tr>
<tr>
<td></td>
<td>P3. Designing feasible experimental procedures</td>
</tr>
<tr>
<td>Analyzing</td>
<td>A1. Conducting an experiment and managing errors</td>
</tr>
<tr>
<td></td>
<td>A2. Using graphs or tables to make sense of data</td>
</tr>
<tr>
<td></td>
<td>A3. Identifying the patterns in data</td>
</tr>
<tr>
<td></td>
<td>A4. Using scientific principles or concepts to interpret data</td>
</tr>
<tr>
<td>Modeling</td>
<td>M1. Describing causal relationships between variables</td>
</tr>
<tr>
<td></td>
<td>M2. Constructing a model by synthesizing several relations</td>
</tr>
<tr>
<td></td>
<td>M3. Applying the model to make a correct prediction</td>
</tr>
</tbody>
</table>

The design of the learning modules

Three learning modules to promote students’ inquiry abilities: Seasons, Environmental issues and Air pollution, were developed based on three overarching design aspects presented in the previous section. In general, teachers spent 10–12 class periods to complete a learning module. Each learning module was composed of four or five activities, and each activity contained several tasks. It should be noted that each task was designed to promote simply one of the 12 inquiry sub-abilities (see Table 2).
### Seasons module

The guiding question for the seasons module was “what are the factors that cause the seasons” (Hsu, 2008). This module consisted of five learning activities. In the first activity, a series of animations about the four seasons was used to engage students in the topic and recall their prior knowledge of seasons. After initial exploration, students were introduced to a computer simulation called SeasonSim (see Appendix 1), in which students can test different variables that may cause the seasons (e.g., latitude, longitude, and the tilting of the earth’s rotation axis). In the second and the third activities, students used SeasonSim to explore how the two factors: latitude and eccentricity, affect the seasons on earth and on an unknown planet respectively. They were required to change the values of latitude and eccentricity, analyze the results produced by SeasonSim, and provide appropriate explanations accordingly. The fourth activity drew students’ attention on the effects of the third factor: the tilting of the earth’s rotation axis, on the changing of the seasons. The fourth activity provided fewer prompts and less guidance, so that students needed to apply what they had learned and work out their own investigations independently. The final activity asked students to create a model that integrated the results from previous activities. Valid data was provided for students to test their models and to weigh the effects of these variables.

### Environmental-issues module

The environmental-issues module consisted of four activities. Geographic information system (GIS) software, My World GIS (http://www.myworldgis.org/), was adopted to support students’ investigations. The teacher first demonstrated how to use the GIS software, and introduced the first task “where to build a train track across America.” Students were guided step by step to build criteria, apply the software to make queries, analyze data, and make a decision. The second activity focused on environmental issues around the construction of a water reservoir. Students searched for valid information and developed their own criteria for selecting the best location (from six potential locations) to build a reservoir. In the third activity, the students used the GIS software to analyze the data based on their own criteria developed in the second activity. Supported by the analysis results, they selected two candidate reservoir locations. The final decision was made based on an evaluation of the pros and cons of these two options. The final activity was the most open-ended. The students were asked to make their own plan on finding a suitable place to build a wind-powered generator on their school campus. They used sensors to collect weather data on the campus and imported them into the GIS software. After analyzing and evaluating the data based on their criteria, they made a decision and presented their plans and decisions to the whole class.

### Air-pollution module

The air-pollution module was designed to help students develop a systematic view about air quality through a series of activities and investigations (Wu, Wu, Zhang, & Hsu, 2013). The first activity engaged students in the issues of air pollution by presenting news and initiating debated questions. Students then worked in groups and used web search engines to explore relevant ideas such as the PSI (Pollutant Standards Index), air quality reporting systems, and air quality testing. By means of the Air-Pollution Modeling Tool (APoMT) software (see Appendix 2), the second and third activities introduced factors that influence the dispersion of air pollutants, such as atmospheric stability, wind speed, and topographic features. Student groups were encouraged to explore the relationships between these factors and develop models to predict air quality.

---

#### Table 2. The modules and the inquiry abilities promoted.

<table>
<thead>
<tr>
<th>Module/Inquiry ability</th>
<th>Questioning</th>
<th>Planning</th>
<th>Analyzing</th>
<th>Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasons</td>
<td>Q2×2</td>
<td>P1</td>
<td>A1</td>
<td>M1×3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3</td>
<td>A3×3</td>
<td>M2×4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4×3</td>
<td>M3</td>
</tr>
<tr>
<td>Environment issues</td>
<td>Q2×3</td>
<td></td>
<td>A1×3</td>
<td>M3×2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A2×2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3×4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4×2</td>
<td></td>
</tr>
<tr>
<td>Air pollution</td>
<td>Q2×4</td>
<td>P1×2</td>
<td>A2</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2</td>
<td>A3×4</td>
<td>M2</td>
</tr>
</tbody>
</table>
variables and the concentration of air pollutants. They needed to establish logical relationships and used them to explain the reasons why the air quality improves/worsens in some locations or under certain weather conditions. In activities four and five, the student groups used the APoMT software to build a model of air pollutant dispersion. They were required to use this model to predict the air quality in two given scenarios, and evaluated the effectiveness of their models. Based on the model, the students created a set of criteria for the next task: where is the best location to build a thermal power plant.

**Methods**

**Participants and settings**

This two-year study was conducted at a public secondary school in southern Taiwan. The school was located in an industrial district with a sound academic reputation. Most of the students were from blue-collar families with a wide range of educational backgrounds. A total of 51 seventh graders (13-14 years old) in two classes participated in the study. These two classes were randomly assigned to the inquiry and the baseline group: 24 in the inquiry group (9 males and 15 females) and 27 in the baseline group (14 males and 13 females). Both classes were taught by an experienced science teacher who held a master’s degree in Earth Science and had taught Science for 20 years. The baseline group was taught with conventional instruction methods in terms of lectures; basically, the teacher followed the instructional sequence given by a textbook. In addition to conventional instruction, the inquiry group was engaged in three technology-infused learning modules during the two-year period. Figure 1 shows an overview of the implementation sequence of the three learning modules and assessments.

![Figure 1. An overview of the implementation of the learning modules and assessments](image)

**Assessment tools**

Two inquiry ability tests included two parts focusing on different sets of inquiry abilities: a paper-and-pencil test and a performance test (see Table 3). The test items were reviewed and validated by a group of specialists, including three university professors, four graduate students, and two science teachers. Both parts were used for the first, second, and third tests.

<table>
<thead>
<tr>
<th>Questioning</th>
<th>Planning</th>
<th>Analyzing</th>
<th>Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>A3</td>
<td>M1×2</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>A2</td>
<td>M2</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>A4×2</td>
<td>M4</td>
<td></td>
</tr>
</tbody>
</table>

**Paper-and-pencil test**

The paper-and-pencil test (see Appendix 3) comprised a total of 13 questions organized into three sets. Some of the questions were created by the authors, while others were modified questions from Lawson and Warren (1976) and
from the TIMSS 2003 Special Initiative in Problem Solving and Inquiry (International Association for the Evaluation of Educational Achievement, 2005). In general, students took approximately 45 minutes (one class period) to complete the test.

**Performance test**

Since it was difficult to assess some of the inquiry abilities through paper-and-pencil testing (e.g., P2: controlling and manipulating variables effectively), we adopted a performance test. The topic of the test was the motion of a cart on a ramp. Provided with equipment and materials, students were required to design and carry out two experiments to test out how different variables might affect the speed of the cart. Specifically, they were required to identify independent and dependent variables in their experiments, design experimental procedures, use equipment to collect data, analyze the data by drawing graphs, and generate conclusions (see Appendix 3).

**Data analysis**

Students' performance in the tests was analyzed according to the scoring rubrics (Appendix 2). The rubrics were developed based on the definitions of the inquiry abilities and the students' responses. Three levels were used (scores of 0, 1, and 2) to categorize the students' performance in each learning task (i.e., one specific inquiry ability). The rubrics were reviewed and revised several times until they best reflected the students' performance in the tests. The reliability of the rubrics was checked by having two researchers analyze the same set of data. The interrater agreement for the inquiry test was .95. In order to identify the students' best-improved inquiry abilities, two-way ANOVA was implemented to compare the differences between the inquiry and baseline groups for the three measurements. For the results that had interaction between teaching conditions and inquiry measurements, the simple effect was tested, while the main effect test was used when the ANOVA results showed no interactions.

The design features were identified by scrutinizing the instructional components of the tasks for which the students demonstrated significant improvements. First, the tasks that involved the students' most-improved inquiry abilities were extracted from the three modules. For example, given that A4 was one of the most significantly improved abilities, five tasks were extracted from the seasons and environmental issue modules. The constant comparative method (Corbin & Strauss, 2008) was then used to analyze the combinations of the instructional components of all of the students' most-improved inquiry abilities. Three main categories emerged from the analysis: (1) The complexity of the task (2) The functions of the prompts (3) The types of the representations presented in the task (see Table 5). These three categories then were used to code the features of the learning tasks that the students improved significantly.

**Results and discussion**

This section first reports the results of two-way ANOVA and the significantly improved inquiry abilities. We then summarize the design features of the tasks that the students improved the most, and discuss how the identification of the features may provide insights into future design in computer-based inquiry environments.

**The inquiry abilities that improved significantly**

A two-way ANOVA test was adopted to examine the effects of the two factors: teaching conditions (inquiry vs. baseline) and inquiry measurements (repeated three times). Due to the small sample and the ordinal scale of measures used in the tests, the Z scores for all of the scores from all three tests were calculated. The results of the two-way ANOVA are presented in Table 4. Overall, it was found that the inquiry group significantly outperformed the baseline group in five inquiry abilities: P1, P3, A4, M1, and M2.

The ANOVA results indicate that the interaction between teaching conditions and inquiry measurements significantly influenced the students' performance on M2 ($F(2,45) = 4.62, p < .05$). Therefore, the simple effect between the two factors was tested. The results indicated that the inquiry group performed M2 ($F(1,45) = 8.54, p < .01$) significantly
better than the baseline group only in the second test. This suggests that the M2 tasks in the Seasons module were effective in helping the inquiry group to construct a model by synthesizing several relationships. For P1, P3, A4, and M1, there were no significant interaction effects between teaching conditions and inquiry measurements. The teaching condition (i.e., inquiry or baseline) was the main factor influencing the students’ development of inquiry abilities. These group differences indicated that relevant tasks across the three modules significantly improved the inquiry-group’s abilities to design experiments (P1) and feasible procedures (P3), use scientific knowledge to interpret data (A4), and describe causal relations between variables (M1).

Table 4. Results of two-way ANOVA

<table>
<thead>
<tr>
<th></th>
<th>Baseline group</th>
<th>Inquiry group</th>
<th>ANOVA results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st M</td>
<td>SD</td>
<td>2nd M</td>
</tr>
<tr>
<td>Q1</td>
<td>1.22</td>
<td>.70</td>
<td>1.22</td>
</tr>
<tr>
<td>Q2</td>
<td>1.67</td>
<td>.55</td>
<td>1.78</td>
</tr>
<tr>
<td>P1</td>
<td>1.07</td>
<td>.87</td>
<td>1.15</td>
</tr>
<tr>
<td>P2</td>
<td>.85</td>
<td>.46</td>
<td>.96</td>
</tr>
<tr>
<td>P3</td>
<td>1.26</td>
<td>.86</td>
<td>1.44</td>
</tr>
<tr>
<td>A1</td>
<td>.85</td>
<td>.60</td>
<td>.70</td>
</tr>
<tr>
<td>A2</td>
<td>1.26</td>
<td>.81</td>
<td>1.48</td>
</tr>
<tr>
<td>A3</td>
<td>.74</td>
<td>.90</td>
<td>.93</td>
</tr>
<tr>
<td>A4</td>
<td>.48</td>
<td>.58</td>
<td>.89</td>
</tr>
<tr>
<td>M1</td>
<td>1.19</td>
<td>.96</td>
<td>1.63</td>
</tr>
<tr>
<td>M2</td>
<td>.48</td>
<td>.51</td>
<td>.37</td>
</tr>
<tr>
<td>M3</td>
<td>.81</td>
<td>.68</td>
<td>.67</td>
</tr>
</tbody>
</table>

Note. *p < .05.

The mean and SD values are the original scores from the inquiry tests; the Z scores are for the two-way ANOVA.

The design features that might contribute to the progress of P1, A4, M1 and M2

Table 5~ Table 7 summarizes the design features of the tasks that students significantly improved in terms of the three coding categories: (1) The complexity of the task: the number of variables involved, (2) The functions of the prompts: clarifying the task goal or providing direction to complete the task, (3) The types of the representations presented in the task: text, table, diagram or animation (note that P3 was not included in the feature analysis as there was only one task across the modules).

The complexity of the task: From simple to complex

Table 5 shows that decomposing multi-variable problems into several single-variable tasks (CT1 and CT2) appeared to be very useful for supporting students designing practicable experiments with relevant variables (P1), using
scientific concepts to interpret data (A4), and describing causal relationships between variables (M1). In fact, the abilities of P1, A4 and M1 were connected as they were concerned with the effects of variables and their interwoven relationships. Therefore, considering only one variable at a time could help focus students’ attention on this particular variable and thus make the tasks simpler for students. Indeed, students often feel confused and do not know where to start when they encounter a learning task that involved many variables or complex relationships. Complexity in learning environments creates a heavy cognitive load for students. Using decompositions to set useful boundaries for learners can restrict the space of activities and thus decrease the complexity (Quintana et al., 2004). Providing structures for complex learning tasks can decrease the cognitive load to a more manageable level. This notion echoes previous research that for complex topics, learning tasks need to be structured in a way that lessens the learners’ cognitive loads (Quintana et al., 2004; Reiser, 2004).

Table 5. Complexity analysis of P1, A4, M1 and M2 tasks

<table>
<thead>
<tr>
<th>Season</th>
<th>Environment issue</th>
<th>Air pollution</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>CT2</td>
<td>CT1×2→CT1×2</td>
</tr>
<tr>
<td>A4</td>
<td>CT1×3→CT1×3→CT2</td>
<td>CT2→CT2</td>
</tr>
<tr>
<td>M1</td>
<td>CT2→CT2→CT2</td>
<td>N</td>
</tr>
<tr>
<td>M2</td>
<td>CT2→CT2→CT2→CT3</td>
<td>CT2×2</td>
</tr>
</tbody>
</table>

Note. CT = The complexity of the task; CT1 = within one factor; CT2 = single factor; CT3 = synthesis of multiple factors. N = No task in the module involved the inquiry ability. “→” is used to separate two items in the same task.

According to the ANOVA results, students’ M2 ability improved significantly only after the seasons module. It is also noted that among the three modules, only the Seasons module posed an integrated question in the final activity that required students to synthesize previous results (CT3). Therefore, we suspected that the structure of the M2 tasks in the seasons module, posing an integrated question in the final activity in particular, had direct impacts on students’ M2 ability. Although the integrated question was more difficult as multi variables were involved, it provided opportunities for students to consider and weigh the influences caused by multiple factors, and which enhanced the students’ ability to synthesize several relations and construct a model (M2). As Reiser (2004) suggested, making some aspects of students’ work more “problematic” is also an important mechanism in supporting students’ learning in computer-based inquiry environments. Directing students to encounter more difficult tasks presents challenges in a short term, but this may trigger productive learning in a long term view. In sum, for better organizing a sequence of learning tasks, designers may follow the principle—from simple to complex. Single-variable tasks can be arranged in early stage so that students can gradually build on their knowledge. More difficult tasks that involve multiple variables can be arranged later for helping integrate relevant concepts and deal with complicated tasks.

Providing prompts: Increasing or decreasing

Table 6 shows that most of the functions of the prompts were actually very similar, that is, to direct students to complete the task. However, different inquiry abilities seemed to show different trends in increasing or decreasing the number of the prompts provided. Regarding P1, M1 and M2, the number of prompts appeared to increase either across tasks within one module or across the three modules. This indicated that an increase in the number of prompts helped students to master complex tasks. In other words, the more complex the tasks were, the more the prompts were needed for supporting students to design practicable experiments (P1), describe causal relationships between variables (M1), and construct a model by synthesizing several relations (M2). As we discussed previously, when encountered the tasks with multiple variables, students needed more supports guiding them to apply what they have learned to this more complex situation.

Table 6. Analysis of the function of the prompts in P1, A4, M1 and M2 tasks

<table>
<thead>
<tr>
<th>Season</th>
<th>Environment issue</th>
<th>Air pollution</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>PR2</td>
<td>PR2×2→PR2×2</td>
</tr>
<tr>
<td>A4</td>
<td>PR2×3→PR2×3→N</td>
<td>PR2→PR2</td>
</tr>
<tr>
<td>M1</td>
<td>N</td>
<td>PR2×2</td>
</tr>
<tr>
<td>M2</td>
<td>→→→→→PR1,PR2</td>
<td>PR2×2</td>
</tr>
</tbody>
</table>

Note. PR = The function of the prompts; PR1 = clarify the task goal; PR2 = provide direction to complete the task. “→” = The task does not provide prompts.
In contrast to P1, M1, and M2, the analysis of the ability of using scientific principles and concepts to interpret data (A4) shows a decrease in the number of prompts across the modules. In the seasons module, fill-in-the-blank questions with more embedded guidance were used first for helping learners to identify the relationship between latitude and seasonal temperature variations. Open-ended questions were posed later with some written prompts providing supports to develop explanations for the relationships between variables. In the last steps, students were asked to construct a model that considered all of the major variables mentioned in previous activities, without any guidance and support. The analysis seemed to suggest that the activities with less support offered opportunities for learners to anchor the abilities just being trained, and had them to apply the abilities to a new context. The activities with the least support allowed students to perform self-regulated inquiry abilities. Such a design is in accordance with the essence of scaffolding that operates a continuum from modeling to coaching and then to fading. More specifically, a design with more prompts and hints at the beginning and more open-ended questions at the end implies a gradual transfer of responsibility to the students (McNeill et al., 2006; Stone, 1998). When the scaffolds gradually fade, learners are expected to take more responsibility in the learning tasks and to implement what they have learned.

Types of the representations in the tasks

The analysis of the representations (see Table 7) shows that these learning tasks were presented in a variety of forms as combinations of texts, tables, diagrams, and animations. In the Seasons modules, students were asked to compare the outcomes from the visualization tools under different conditions (the amount of solar energy at the perihelion and the aphelion in the seasons module). Relevant data was presented in different forms simultaneously, in this case, diagrams and tables. The use of tables together with diagrams drew students’ attention to the patterns and the changing between the two variables, and which may greatly support them to use relevant scientific principles and concepts to interpret data (A4). Similarly, in the Air pollution module, students were asked to compare the effects of different variables on the dispersion of pollutants through the visualization tools. It was found that using more than one type of diagrams, bird-view and cross section diagram in this case, helped students gain deeper understanding of the effects of each variable, which in turn may provide solid foundation to describe causal relationships between variables (M1).

Table 7. Analysis of the types of the representations presented in P1, A4, M1 and M2 tasks

<table>
<thead>
<tr>
<th>Season</th>
<th>Environment issue</th>
<th>Air pollution</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>MR1, MR2, MR3, MR5→</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>MR1, MR3, MR5→</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MR1, MR3, MR5</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>MR1→MR1→MR1→</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>MR1→MR1→MR1→</td>
<td></td>
</tr>
</tbody>
</table>

Note. MR = The types of representations; MR1 = text; MR2 = table; MR3 = diagram; MR4 = more than one type of diagram; MR5 = animation.

The findings suggest that when the content is presented by synergistic representations, students can benefit from multiple and co-occurring forms of curricular support types (McNeill & Krajcik, 2009; Tabak, 2004). Synergistic representations should align with a defined learning goal. That is, representations and scientific processes need to be paired based on the affordance of the representations and the goals of the activities (Wu & Puntambekar, 2012). In addition, using multiple representations may have greater potential to help diverse learners with different levels of background knowledge and abilities in a classroom. Therefore, to meet students’ needs, instructional designers need to assess students’ prior knowledge before setting up an appropriate learning goal. Based on the learning goal, designers then select several representations to display the content and organize them structurally and synergistically.

Conclusion

This two-year study examined the effects of a series of technology-infused learning modules on the students’ development of inquiry abilities. Through the lens of students’ learning outcomes, we identified some design features that potentially enhanced students’ ability to manipulate variables, reason the relationships between variables, and
incorporate scientific concepts to interpret data and construct explanations. For a complicated learning context, the activities need to be organized from simple to complex, so that learners are prepared to clarify and deal with the complexity step by step. When learning tasks involved multiple variables, relevant data or information can be presented with more than one representation type. Synergistic representations help draw learners’ attention on the trends and patterns in data and gain deeper understanding of causal relationships. It is worth noting that for different inquiry abilities, designers may use different strategies to present scaffolds. For some inquiry abilities such as using scientific principles and concepts to interpret data, it might be more effective to decrease the supports over time because more responsibilities should be given when learners become more capable. However, for complex tasks involved multiple variables at a later learning stage, students may still need certain prompts guiding them to synthesize the relationships between variables and construct a model; even they have learned each variable before. The design features identified in the study provide some useful guidelines for teachers and curriculum designers to design learning tasks for different inquiry abilities.

References


Appendix 1: Screenshots of the software used for learning modules

(1) Screenshot of SeasonSim
(2) Screenshot of My world GIS (http://www.myworldgis.org/)

(3) Screenshot of Air-Pollution Modeling Tool
## Appendix 2: Scoring rubrics for inquiry abilities

<table>
<thead>
<tr>
<th>Inquiry abilities</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Questioning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1. Posing a testable question</td>
<td>2</td>
<td>Posing a testable question that is relevant to a scenario and involves relationships between variables.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Posing a testable and descriptive question that is relevant to a scenario.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Posing a question that is untestable or irrelevant to a scenario.</td>
</tr>
<tr>
<td>Q2. Recognizing key variables of a question</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Recognizing key variables of a question and indicating relationships between the variables.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Recognizing key variables of a question, but not identifying relationships.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Not recognizing variables of a question.</td>
</tr>
<tr>
<td><strong>Planning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1. Designing an experiment with relevant variables</td>
<td>2</td>
<td>Designing an experiment that can verify all relationships between the key variables.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Designing an experiment that can verify a few of the relationships between the key variables.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Not being able to design an experiment, or designing an experiment that cannot verify any relationships.</td>
</tr>
<tr>
<td>P2. Controlling and manipulating variables effectively</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Controlling and manipulating several variables correctly and simultaneously.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Controlling and manipulating one variable correctly.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Controlling incorrect variables.</td>
</tr>
<tr>
<td>P3. Designing feasible experimental procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Designing feasible experimental procedures that can answer the research question.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Designing feasible experimental procedures, but these procedures cannot answer the research question.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Designing experimental procedures that are not feasible.</td>
</tr>
<tr>
<td><strong>Analyzing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Conducting an experiment but ignoring experimental errors.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Not being able to conduct an experiment.</td>
</tr>
<tr>
<td>A2. Using graphs or tables to make sense of data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Using meaningful graphs or tables to make sense of data and correctly labeling data units.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Using graphs/tables to represent data, but these data do not represent meaningful relationships; or not correctly labeling data units.</td>
</tr>
<tr>
<td><strong>Modeling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1. Describing causal relationships between variables</td>
<td>2</td>
<td>Identifying variables and describing causal relationships between them.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Identifying variables without a description of relationships between them.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Identifying irrelevant variables or incorrect relationships.</td>
</tr>
<tr>
<td>M2. Constructing a model by synthesizing several relations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Constructing a model to interpret results and comparing the effects of two variables.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Constructing a model to interpret results, but not being able to compare the effects of two variables.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Identifying the effect of one variable or constructing a model that</td>
</tr>
<tr>
<td>M3. Applying the model to make a correct prediction</td>
<td>2</td>
<td>Applying the model to make a correct prediction and providing reasons.</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>----</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Applying the model to make a correct prediction but not being able to provide any reason.</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Not being able to apply the model to make a prediction.</td>
</tr>
</tbody>
</table>
Appendix 3: Sample items of the pencil-and-paper test and the performance test

Paper-and-pencil test

Set 1: Why do scooter exhaust pipes get rusty?
David lives in the Cian-Jhen area of Kaohsiung city. He left his scooter outdoors and found that the exhaust pipe got rusty. One day, he went to visit his aunt around the Port of Kaohsiung. He found that his aunt’s scooter exhaust pipe was more seriously rusty than his, even though the two scooters were exactly the same, left outdoors, and bought at the same time.

1. If you would like to conduct a scientific experiment with regard to the scenario stated above, what is your research question? (Please write down a research question that can be tested by a scientific experiment.)

2. Please predict possible results and explain the reasons that support your prediction.

3. Regarding the research question you posed in question 1, what are the variables involved in the experiment?

Please write down the experimental procedures.
Performance test

The motion of a cart on a ramp

The aim of the experiment is to understand what the factors are to affect the cart speed. For each experiment (two in total), you need to pose a research question, describe the research design, and use the data you collect to present the results and answer the research question.

With reference to the device presented above, please use the equipment provided (ramp, cart, etc.) to conduct two experiments, and answer the following questions. After you complete the experiments, please conclude how cart speed is affected by the variables.

A: Research question 1:

1. Research Design (Table 1)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control variables</th>
<th>Independent variable</th>
<th>Dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental procedures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>steps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicting results</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Laboratory records: (Please use appropriate tables to record data.)

3. Please use appropriate charts (line chart, bar chart, … etc.) to present data.

4. Please describe the experimental results based on the chart presented above.

5. Do the experimental results prove your prediction (in Table 1) correct? If not, what might be the reasons?

6. Answer the research question 1 based on the experimental results.

(The table and questions are the same in A: Research question 1 as in B: Research question 2)

Conclusion: Sum up the results of the two experiments, your conclusion is: ___
A Social Learning Management System Supporting Feedback for Incorrect Answers based on Social Network Services

Jiseong Son¹, Jeong-Dong Kim²*, Hong-Seok Na³ and Doo-Kwon Baik¹*

¹Department of Computer and Radio Communications Engineering, Korea University, Seoul, Korea // ²Department of Computer Engineering, Sunmoon University, Asan, Korea // ³Department of Computer and Information Communication, The Cyber University of Korea, Seoul, Korea // redfunky07@korea.ac.kr // kjdvhu@gmail.com // hsnan99@cyberkorea.ac.kr // baikdk@korea.ac.kr

*Corresponding authors

(Submitted January 27, 2015; Revised July 7, 2015; Accepted August 11, 2015)

ABSTRACT

In this research, we propose a Social Learning Management System (SLMS) enabling real-time and reliable feedback for incorrect answers by learners using a social network service (SNS). The proposed system increases the accuracy of learners’ assessment results by using a confidence scale and a variety of social feedback that is created and shared through learners’ personal SNS connections. The system measures the intimacy between learners and their SNS friends, as well as the registered feedback time for reliable social feedback. We examined the performance of the proposed SLMS in relation to collaborative learning through the measurement of learners’ level of satisfaction.

Keywords

Collaborative learning, Learning management system, Social learning, Social network services, Social feedback

Introduction

With the development of the Internet, a considerable amount of research has been carried out on Web-based e-learning, which enables learners to study anytime and anywhere without traditional limitations of place and time (Horton, 2000). In addition, a variety of learning management systems (LMSs) has been developed for the systematic management of courses, grades, progress, and attendance of the students. Representative examples of LMSs include Moodle (Cole & Foster, 2007; Kumar, Gankotiya, & Dutta, 2011), Desire2Learn (D2L) (Wang, & Shao, 2008), and Blackboard (Martin, 2008; Pishva, Nishantha, & Dang, 2010). LMSs support and manage learning content allowing various file formats and media. One of the various features of LMSs is the ability to conduct assessment by providing learners with a quiz (or tests) to evaluate their ability, analyze their correct or incorrect answers, and provide feedback on their incorrect answers. The quiz helps learners recognize the information they did not understand during the learning period. An educator can evaluate a learner’s grade by using learners’ quiz results. In particular, feedback, including an explanation of incorrect answers, helps learners improve their understanding of the learning contents and relearn why their answers were incorrect (Thoms, 2011).

However, the quiz and feedback of existing LMSs has the following limitations: a lack of analysis of whether the learners understand the content; static explanations despite the different learning abilities of the respective learners; and restricted interaction and collaboration feedback among learners. First, existing LMSs lack the ability to analyze whether learners understand the precise meaning of a question. Such systems simply measure the level of understanding and insight of learners by depending on the outcome of analyzing their correct or incorrect quiz answers. Some correct answers may have been given by chance, but the learners are not provided with an explanation and feedback for such answers. To improve the learner’s understanding of the subject material and evaluate their level of understanding and insight, an LMS system needs to be able to analyze whether the learner correctly understands the questions and answers of a quiz. In addition, existing LMSs only provide a previously defined static explanation of an incorrect answer without analyzing the various reasons why the learners may have answered a question incorrectly. Analyzing the various causes for incorrect answers and providing a corresponding explanation require considerable time and cost. To solve incorrect answers, the learners study and acquire additional information from the internet or the course lectures, or ask the educator or someone despite providing basic static explanation. This additional inconvenient process is frequently repeated. As a result, refinement and further development of LMSs is required for efficiently supporting and managing various analyses based upon the causes of incorrect answers and the different levels of learner achievement. Finally, as a general course-based approach to e-learning, such systems have restricted interaction and collaboration feedback among learners (Dalsgaard, 2007).
When a learner has a question related to a course lecture, the learner may ask the educator during the course. The educator can provide feedback to the learners through different questions and lead the students through the course. On the other hand, the relationship among learners is temporary and restricted to the duration of the course. After the end of the course period, the learners may not have the opportunity to ask questions related to their learning. As a result, these limitations make an LMS inappropriate for supporting collaborative learning.

Various researches in education investigating interaction, participation, and sharing among learners have been recently conducted with the evolution of Web 2.0 and the advent of e-learning 2.0 (Yang & Chen, 2008; Chen, Kinshuk, Wei, & Yang, 2008; Huang, Jeng, & Huang 2009). Social media, i.e., Web 2.0 tools (e.g., blogs, wikis, social bookmarking sites, social networking services, and etc.), help users create content that can be shared with other users. When social media is applied to e-learning, the key feature of social software supports the learner in creating and sharing their learning content (Mcloughlin & Lee, 2010; Ford, Bowden & Beard, 2011). The learner therefore becomes an active participant in the LMS course. The relationships among learners support interactive and collaborative feedback. Among the different types of social media, video, pod-casts, wikis, and so on are valuable tools for teaching, and SNS can be valuable tools for collaborative learning (Moran, Seaman & Tinti-kane, 2011). According to a Pearson’s survey, more than 90% of the faculty members in US universities employ social media for personal purpose or professional education, and among them, visits and postings to SNS make up the highest of such usage (Moran, Seaman & Tinti-kane, 2011). SNS-based learning enables the learners to carry out collaborative learning by forming close personal connections on the basis of social learning theory and Connectivism (Siemens, 2005) using their real-world personal connections and SNS based on online social networking (Huang, Yang, Huang, & Hsiao, 2010; Haythornthwaite, & de Laat, 2010). SNS can help learners share information and their opinions for learning with their friends in real time. Teaching and learning utilizing social media have increased continuously. Furthermore, SNS in the field of learning has become more important owing to the social interaction with learners. The use of SNS in learning involves the advantage of real-time, information sharing, simple posting and reliable feedback from friends (Du, Fu, Zhao, Liu & Liu, 2013; Yin, Tabata & Ogata, 2009; Popescu, 2014).

In order to benefit from social networks, we propose a Social Learning Management System (SLMS) that enables real-time and reliable feedback on incorrect answers by incorporating a SNS. The SLMS allows learners to share the necessary information for learning with each other and receive feedback from their personal connections through an SNS. Collaborative learning among learners using SNS minimizes the time and expense required for an analysis of their incorrect answers and provides various and abundant feedback by allowing questions to be shared and analyzed, as well as providing reasons for incorrect answers. In addition, to provide feedback based on a precise evaluation of the learner’s ability, for this study, we measure the learner’s confidence in their answers by expanding the Confidence-Based Marking (CBM) feature proposed in Moodle. This feature is used to evaluate the learner’s ability (Gardner-Medwin, & Curtin, 2007). When compared to existing LMSs, the SLMS can improve learners’ learning ability by supporting feedback on incorrect answers through collaborative learning by relying on the social connections among learners.

Quiz and feedback process of existing representative online LMSs

Moodle is an open source course management system (Cole & Foster, 2007; Kumar, Gankotiya, & Dutta, 2011) that provides an array of tools for educators to manage and promote teaching and learning. Among the various quiz options that Moodle provides, there is an option for offering CBM, analyzing whether the learners have filled out the correct answers with confidence, and feedback options providing an analysis and explanation of the incorrect answers by analyzing the learning results (Gardner-Medwin, & Curtin, 2007). CBM uses a 3-point scale. The level of the learner is determined by granting a credit or penalty based on the confidence level in their answers. In this way, Moodle identifies the exact ability of the learners. However, Moodle does not provide feedback on incorrect answers by utilizing CBM. Moodle provides learners with three types of feedback, namely, overall feedback, general feedback, and specific feedback. Overall feedback is provided to the learners after they submit a quiz; it is dependent on the grade that the learners achieve. General feedback includes some background knowledge about the questions; it too is provided after the quiz is completed; this feedback gets entered for each quiz item by the educator when the quiz is first created. Depending on the learner’s answers, preset specific feedback is also provided. Such feedback is provided by Moodle for learners focusing on preset feedback; the system provides only predetermined static
feedback information without any analysis of the cause of the incorrect answers. It does not provide collaborative feedback between the learners. Blackboard is commercial LMS (Martin, 2008; Pishva, Nishantha, & Dang, 2010) that provides test and test feedback options for determining the ability of the learners. The test feedback option determines the type of feedback provided to the learners after the completion of the test. There are four types of feedback: score, submitted answers, correct answers, and feedback. The score option shows only the learner's test score, and the submitted answers option shows the content of the questions of the test and the answers that the learners selected. The correct answers option shows only the content of the test questions and the correct answers. In addition, it distinguishes between the correct and incorrect answers and doesn’t display the answers chosen by the learners. The feedback option provides an explanation of the incorrect answers. However, the test feedback options don’t consider whether the learner has a good understanding of the questions, and provide only a brief preconfigured explanation.

D2L provides an online environment for creating and managing courses (Wang, & Shao, 2008). The evaluation method of D2L relies on quizzes, discussions, grades, surveys, and drop-boxes. Among them, the quizzes are used to assess the learners by offering various test formats. The difference between other LMSs and D2L is that D2L supports social networking among learners. Learners register and share their profiles through social media in order to connect their various SNS accounts to their educational, personal, and professional communities. However, social networking supported by the current version of D2L shares only the learner’s experiences, goals, and progress with their peripheral personal connections; it does not support any learning through these connections.

Social learning management system based on SNS

System architecture

As shown in Figure 1, the architecture of the proposed system is divided into the SLMS and Social Learning Learner Interface (SLLI). The SLMS consists of a quiz items management module, an incorrect answer management module, a learner management module, and a Social Feedback module.

![Figure 1. System architecture](image)

The SLLI has a series of processes that provide learners with the requested quiz items and allows them to exchange feedback about their incorrect answers. It consists of 5 steps. In the first step, the learner sends a quiz request query to the system and receives the items. In step 2, the learner’s answers to the items and their confidence scale are
evaluated. The confidence scale classifies whether the learner knows the correct answer of an item based on the results analyzed on a 3-point scale that the learner choose when answering quiz items. If the confidence scale is “assurance” and the item is also the correct answer, the algorithm calculates the percentage of correct answers. On the other hand, if the confidence scale is “assurance” but the item is incorrectly answered, the algorithm provides an explanation for the incorrect answer and executes the social feedback module. Similarly, if the confidence scale is “uncertainty” or “unknowingness,” the learner does not clearly understand the item even though their answer might be correct. The SLMS generates the items given to the learners, their answers, and their confidence scale as a resource (R), and handles them by submitting them to the SLMS. In step 3, the results are sent to the learner after their correct and incorrect answers are classified by the incorrect answer management module along with an analysis of the confidence scale based on R. Step 4 provides an explanation for incorrect answers that is a predefined static explanation in the system. The last step provides social feedback based on the SNS through feedback sharing among the learners. R, i.e., the resources described above, can be defined as follows:

**Definition 1.** A resource that is given to the learner is defined as a triple \((I, A, C)\), where \(I\) denotes the type of items and the set of questions, \(A\) denotes the learner’s results for a set of items, and \(C\) indicates their confidence scale.

**Resource R = (I, A, C)**

The quiz items management module consists of the quiz items manager and the items level manager. The quiz items manager contains the items and manages them based on the learning type; it retrieves the items from the item bank by analyzing the type of items that the learners requested or answered incorrectly. The items level manager manages the items by classifying items based on their level of difficulty and determines the level of difficulty of new items or modifies the level of difficulty of the existing items, based on the average rate of correct answers given by the learners. The item bank consists of a variety of item types, where each type of item has a subset of several different types of items. \(R\), described in Definition 1, consists of pairs of \(items (I)\) and \(answers (A)\) because they share a mapping relationship. The following is a formal definition of item and answer.

**Definition 2.** I consists of an ItemPattern, which is the quiz type, and \(i\), which is a subset of the quiz type. \(A\) denotes the answers corresponding to the items. An item and its answer are composed of the pair \((I, A)\).

**ItemPattern = \{ItemPattern_1, ItemPattern_2, ..., ItemPattern_n\}**

**ItemPattern = \{i_1, i_2, i_3, ..., i_d\}**

In case of \(i_i \in \text{ItemPattern}_1, a_1 \in A_i, \text{ItemPattern}_1 \subseteq I\) and \(A_i \subseteq A\), it is defined as \((i_1, a_1), (\text{ItemPattern}_1, A_1)\)

The incorrect answer management module consists of the manager of explanation for incorrect answer, the analysis of confidence scale, and the evaluation engine. The explanation for incorrect answer manager manages basic static explanations based on the items, the learning type, and the level of difficulty of the items, and provides explanations to the learners. The analysis of confidence scale module classifies the items as either correct or incorrect through an analysis of the confidence scale. Based on the results of this analysis, the evaluation engine provides the social feedback manager with the results of the evaluation determining whether the items were answered incorrectly. The aforementioned confidence scale is an evaluation of the learner’s confidence in their answers and is defined as follows:

**Definition 3.** The confidence scale is an evaluation of the degree of confidence of the correct answers for the learner's learning outcomes. Assurance indicates that the learner understands the correct answer exactly. Uncertainty indicates that the learner is uncertain of a correct answer. Unknowingness indicates that the learner does not know the correct answer.

**Confidence Scale C = \{assurance, uncertainty, unknowingness\}**

The learner management module consists of the learner profile manager and the learner social network manager. Through the learner profile manager, the system can analyze a learner’s profile and the learning information. Through this analysis, the level of the learner is determined and the generation of and change in the learning level are managed. Moreover, by managing the login information of the learners enrolled in the system, the system manages whether the learners receive training on a regular basis, the amount of their studying/learning during a certain period of time, and the degree of increase in their learning level. Through this feature, the learners can manage their own learning schedule and their amount of learning. The learner’s social networking manager manages the information of the personal connections among the learners. In particular, to provide feedback for incorrect answers based on SNS-
based feedback through the sharing of incorrect answers, the system uses the information of the learner’s personal connections. The social feedback module consists of the social feedback engine, the intimacy measurement, and the social feedback manager. The social feedback engine shares quiz items, answers, and explanations for incorrect answers on the various SNS used by the learners. In addition, the social feedback engine ranks the social feedbacks by weighing the intimacy between a learner and their SNS friends as well as the registered time of social feedback. The social feedback is registered by the learner’s SNS friends in real time. When new social feedback is registered, the social feedback engine compares it with previous registered social feedback, and then provides a high ranking to social feedback having a higher value of intimacy and that is more recently registered. Therefore, the SLMS can avoid blind advertisements or spam from strangers.

The intimacy measurement measures the degree of closeness between learners and their SNS friends. Various researches on the measurement of intimacy between people have been conducted (Liang, Li & Turban, 2009; Seol, Kim, Shim & Baik, 2012; Seo, Kim & Baik, 2014). Among such various researches, we consider the ratio of sharing friends as a method for calculating the intimacy among users (Seol, Kim, Shim & Baik, 2012; Seo, Kim & Baik, 2014). The following is a definition of the relationship between learners and their SNS friends.

**Definition 4.** Intimacy \( I \) is an affinity between learner and SNS friend. \( I \) consist of Learner \( L \), SNS Friends \( F \), and Edge of Relationship between learner and SNS friend \( E_d \).

**Intimacy \( I \) = (L, F, Ed)**

\[ I = \{l_1, l_2, \ldots, l_n\} \] is a set of learners.

\[ F = \{F_{l_1}, F_{l_2}, \ldots, F_{l_n}\} \] is a set of SNS friends.

\[ F_{l_1} = \{f_{1,1}, f_{1,2}, \ldots, f_{1,n}\}, \quad F_{l_2} = \{f_{2,1}, f_{2,2}, \ldots, f_{2,n}\}, \ldots, \quad F_{l_n} = \{f_{n,1}, f_{n,2}, \ldots, f_{n,n}\} \] is a set of SNS friends of a learner.

\[ E_d(l, F) \] is the relationship between a learner and their SNS friends.

The social feedback manager manages the registered social feedbacks. The social feedback manager includes the following information for registered social feedback: 1) the type of item, 2) who provided the social feedback, and 3) when the feedback was written. Based on this information, the social feedback engine measures the ranking of the registered social feedback in real-time, and the social feedback manager then reflects the ranking results. The definition and ingredients for social feedback are as follows.

**Definition 5.** Social Feedback \( S \) is feedback for sharing an item written by SNS friends of the learner. \( S \) consists of Resource \( R \), Intimacy \( I \) between the learner and their SNS friends, and the registered time \( T \) of the social feedback.

**Social Feedback \( S \) = (R, I, T)**

**Algorithm for the social feedback**

In this section, we describe an algorithm for processing the social feedback. The algorithm processes to share the information related to an items on SNS and to provide the learners with feedback from their SNS friends. As shown in Algorithm 1, for information on an item, the input is \( R, I \) is the intimacy between the learner and their SNS friends, and \( T \) is the time the feedback is written. The output, \( S \), is the registered social feedback by the learner’s SNS friends. In Lines 1 and 2, \( l_i \) is the learner \((L)\) using the SLMS, \( F_{l_i} \) is the set of SNS friends of \( l_i \). \( i \) is the identification number of the learner. In Lines 4 and 5, if \( l_i \) as a learner has SNS friends \((F_{l_i})\), \( R \) is shared with every \( F_{l_i} \). In lines 6 and 7, social feedback \( S_{l_i,j} \) by written the learner’s \( j \)th SNS friend \((f_{l_i,j})\) consists of \((R_n, I_{l_i,j}, T)\), such as described in Definition 5. The social feedback is affected by the intimacy between the learner and their SNS friends and the registered time of the social feedback. In other words, an SNS friend that has a higher value of intimacy can provide more reliable social feedback. Recent registered social feedback has more reliability than older feedback because a social network is a real-time system and recent registered social feedback refers to previous registered social feedback. As a result, line 8 measures the intimacy between the learner and their SNS friends, and then ranks \( S_{l_i,j} \) by...
reflecting the registered time of the social feedback. Finally, line 9 returns \( S_{i,j} \) reflecting the measured result of the ranking.

**Algorithm 1. Algorithm for Social Feedback**

**INPUT:**
• \( R \), a set of resources \( R = (I, A, C) \)
• \( In \), a set of Intimacy \( In = (L, F, Ed) \)
• \( T \), time the feedback is written

**OUTPUT:** \( S \), Social Feedback

**METHOD:**
1: \( l_i \in L \), \( l_i \) is a Learner
2: \( F_{l_i} = \{ f_{l_i,1}, f_{l_i,2}, \ldots, f_{l_i,k} \} \), \( F_{l_i} \) is a set of SNS friends of \( l_i \)
3: \( i \) is the identification number of the learner
4: if \( l_i \) has \( F_{l_i} \) then
5: Sharing \( R_n \) with \( \forall F_{l_i} \)
6: for \( j = 1 \) to \( k \) do
7: Social_Feedback \( S_{i,j} \leftarrow (R_n, In_{i,j}, T) \)
8: ranking \( S_{i,j} \leftarrow (1 - \lambda) \cdot \left( l_i \leftrightarrow f_{l_i,j} / | F_{l_i} | \right) + (\lambda \cdot t_{i,j}) \)
9: return \( S_{i,j} \)
10: endfor
11: endif

We propose an equation for measuring the ranking of social feedback, shown in (1). In social network, the intimacy is determined by ratio of sharing friends. That is, the greater the number of sharing friends, the higher the intimacy measured. The smaller the number of sharing friends, the lower the intimacy measured.

\[
S_{i,j} \leftarrow (1 - \lambda) \cdot \left( l_i \leftrightarrow f_{l_i,j} / | F_{l_i} | \right) + (\lambda \cdot t_{i,j}) \quad (1)
\]

In (1), \( \lambda \) is a constant value. We decide a \( \lambda \) value of 0.1 for weighting the direct friend relationship than the registered time of the social feedback because the quality of social feedback that helps the learner is more important than when the feedback is given. \( l_i \leftrightarrow f_{l_i,j} \) is the number of sharing friends among the learner \( l_i \) and their SNS friends \( f_{l_i,j} \).

As the registered time of feedback, \( t_{i,j} \) is calculated based on the up-to-the-minute and oldest registered time of social feedback. Recent registered social feedback has higher reliability rather than older feedback because a social network is a real-time system and recent registered social feedback refers to previous registered social feedback. Therefore, \( t_{i,j} \) holds the weight of the recent registered social feedback. The registered time of the feedback is \( t_{feedback} \), and the up-to-the-minute registered time is \( t_{new} \). The previous registered time of the feedback is \( t_{old\_feedback} \). Therefore, we propose (2) for the registered time of social feedback \( t_{i,j} \).

\[
t_{i,j} = 2^{t_{new\_feedback} / t_{new\_old\_feedback} - 2^{-1}} \quad (2)
\]

**Implementation**

We describe the implementation of the SLMS based on the SNS. A quiz that is sent to the learner’s mobile device consists of the items, a 3-point confidence scale, and answer checks, as shown in Figure 2. As shown in Figure 2-(A), the learner selects the answers to all items and submits them to the system along with their level of confidence in their answers. As shown in Figure 2-(B), the correct or incorrect answers are classified by analyzing the submitted answers and the level of confidence. As shown in Figure 2-(C), for items that the learners answer incorrectly, the proposed system provides a basic static explanation registered in the SLMS. The learners can go through the
explanation and correct answers provided by the SLMS. They can determine the reason for their incorrect answers based on the explanations given. However, there is a limit to such an analysis because an explanation provides the same information to all learners. As a result, the learners share their explanation and other details of an item through their SNS, as shown in Figure 3-(A). The learner’s peripheral personal connections on Facebook provide feedback to the learner. As shown in Figure 3-(B), the learner’s SNS friends write their feedback that enables them to verify the type of learning, as well as the items, annotations, answers, explanation, degree of difficulty, etc. A variety of social feedback provided by the learner’s SNS friends can be verified through the explanations of their incorrect answers in the SLMS (Figure 3-(C)).

![Figure 2. (A) An item, (B) the learner’s submitted answer and correct or incorrect answer verification, and (C) the provisioning and sharing of the incorrect answer](image)

![Figure 3. (A) Learner’s SNS newsfeed share, (B) feedback provisioning, and (C) feedback result verification](image)
Experiment and evaluation

Our experiment targeted a specific quiz (TOEIC) and can be categorized as a one-group post-test quasi-experimental design since no random assignments were performed. The purpose of this experiment was to evaluate the SLMS performance in relation to collaborative learning through the measurement of learner’s level of satisfaction. We examined the measurement of learner satisfaction to compare the two different types of systems. One system is SLMS without a confidence scale and social feedback similar to existing LMSs, and the other is SLMS with a confidence scale and social feedback. The experiment was conducted over a period of two months, i.e., April and May of 2013. For the experiment, 258 university students attending Korea University (130 students) and the University of Seoul (128 students) who were attending the same class during the same semester were chosen.

![Figure 4. Basic investigation of the experiment participants](image)

**Basic experiment (Participants)**

The participants were surveyed on their use of SNS and the registration of an online lecture, as shown below.

**Use of SNS**
- 4-(A) Do you use SNS?
- 4-(B) If you do, what SNS are you using now (multiple answers allowed)?
- 4-(C) How frequently do you use the SNS daily?
- 4-(D) How long do you use the SNS daily?

**Registration for online lecture**
- 4-(E) Have you ever taken an online lecture?
- 4-(F) How long did you take the online lecture?
- 4-(G) Were you satisfied with the explanation for incorrect answers provided in the online lecture you took?

As shown in Figure 4, among the 258 participants, 88% (228 persons) answered that they use an SNS (Figure 4-(A)), 53% (234 persons) reported that Facebook is their most-used and most-visited site. 34% (152 persons) reported that they use Cyworld (Figure 4-(B)). Regarding how many times they visit SNS daily, 43% (98 persons) said 1 to 5 times, 21% (49 persons) said 5 to 10 times, 28% (63 persons) said 10 to 30 times, and 8% (18 persons) answered that they visit an SNS more than 30 times (Figure 4-(C)). Regarding how many hours they use SNS daily, 35% (79 persons) said 10 to 30 minutes, which was the highest proportion, followed by 30 minutes to 1 hour at 28% (65 persons), and 5 to 10 minutes at 20% (45 persons) (Figure 4-(D)). Through the above results, the majority of the participants used an SNS each day. Regarding their experience with online lectures, 81% (208 persons) of the participants answered that they had taken an online lecture before (Figure 4-(E)); in addition, 49% (102 persons) of them took an online lecture for more than one year, whereas 26% (54 persons) and 19% (40 persons) took a lecture
for more than six and three months, respectively (Figure 4-(F)). Meanwhile, of the participants who responded to the question regarding their satisfaction of the explanations for incorrect answers provided during their online lecture, 32% (55 persons) answered that they Agree, 36% (61 persons) neither Agree nor Disagree, and 32% (55 persons) Disagree (Figure 4-(G)). Most of the participants responding to the basic experiment had experienced an online lecture (three months to more than one year) and received an explanation for their incorrect answers. They also stated that the explanation for their incorrect answers contributed greatly to their learning improvement, that is, the explanation for incorrect answers had a significant effect on their learning result (Figure 4-(H)).

Through the basic experiment, it was found that most of the participants experienced an online lecture in the past, the explanations for their incorrect answers provided by the online lecture had a significant effect on their learning, and the participants are users of an SNS. Therefore, as an experimental group for the performance experiment of the SLMS, another experiment evaluation was conducted on the 198 students who stated they have experienced an online lecture in the past and use Facebook.

**Experiment**

For the experiment evaluation, which aimed to determine the participants’ level of satisfaction, 198 participants were given TOEIC questions to study using the SLMS for a two-month period. The items used in the experiment were 10 practice tests excerpted from reading comprehension Part 5, 6, and 7 in TOEIC.

For the first month, the participants were instructed to learn the material using the SLMS in the same manner as a traditional online learning method with no confidence scale or social feedback. During the second month, the participants were asked to learn using the SLMS. This time their learning involved with a confidence scale and social feedback. The participants shared the information regarding the TOEIC questions and incorrect answers through their personal Facebook accounts linked with the SLMS. They could thus receive social feedbacks from their close friends. The learner’s SNS friends (i.e., how to solve a problem and knowledge regarding the solution) were registered and managed by the Social Feedback Manager of the SLMS. During the experiment period, the participants could log into their personal accounts registered in the SLMS system using their personal smartphones. This ensured anytime-anywhere learning. After the two-month learning experiment a survey was conducted to evaluate learners’ level of satisfaction with the SLMS system. Results are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Survey results</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey item</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>I. The explanation for incorrect answers reflect a level of learners</td>
<td>0%</td>
<td>8%</td>
<td>15%</td>
<td>57%</td>
<td>40%</td>
</tr>
<tr>
<td>II. The learners can find the reason for their incorrect answers through the provided explanation</td>
<td>0%</td>
<td>15%</td>
<td>26%</td>
<td>49%</td>
<td>35%</td>
</tr>
<tr>
<td>III. The learners can reflect the feedback with regard to their own incorrect answers</td>
<td>1%</td>
<td>13%</td>
<td>14%</td>
<td>59%</td>
<td>37%</td>
</tr>
<tr>
<td>IV. The explanations for incorrect answers can be shared (including previous used online lecture systems)</td>
<td>0%</td>
<td>17%</td>
<td>11%</td>
<td>49%</td>
<td>34%</td>
</tr>
<tr>
<td>V. The system provide sufficient explanations for incorrect answers</td>
<td>0%</td>
<td>15%</td>
<td>14%</td>
<td>45%</td>
<td>35%</td>
</tr>
<tr>
<td>VI. The system with providing explanations for incorrect answers is conducive to the learning effect</td>
<td>0%</td>
<td>16%</td>
<td>16%</td>
<td>50%</td>
<td>36%</td>
</tr>
</tbody>
</table>

*Note. A = SLMS without Confidence Scale & Social Feedback, B = SLMS (Proposed System). 1 = Strongly Agree, 2 = Agree, 3 = Agree nor Disagree, 4 = Disagree, 5 = Strongly Disagree.*

**Experiment result and evaluation**

During the two-month period a total of 1,000 items were provided to the participants. On average, the participants had to learn the answers to 352 items. In addition, 942 items (including the duplicate items) were shared through the SNS, which indicates a sharing of 4.75 items per participant during the experiment. Furthermore, the total number of
social feedbacks received with regard to the shared items was 3,436. This represents 3.64 feedbacks per shared item on average. Table 2 shows the detailed statistics of the experiment results.

Table 2. SLMS usage statistics

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Total</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td># Participants</td>
<td>198</td>
<td>-</td>
</tr>
<tr>
<td># Quiz (items)</td>
<td>1000</td>
<td>352</td>
</tr>
<tr>
<td># Sharing</td>
<td>942</td>
<td>4.75</td>
</tr>
<tr>
<td># Social Feedback</td>
<td>3436</td>
<td>3.64</td>
</tr>
</tbody>
</table>

Table 1 shows the survey result of the satisfaction level after using the two different types of systems. Table 1-A shows the satisfaction level with regard to the system in which no confidence scale and social feedback were provided. The response to the question of whether the explanation for an incorrect answer reflected the level of the learner revealed that 45% of the respondents believe it did not, whereas 15% believe it did. The response to the question of whether learners can find the reason for their incorrect answers through the provided explanation indicated that 39% said they cannot find the reason while 26% answered yes. Regarding the question of whether learners can reflect the feedback about their incorrect answers, 48% did not agree that they can reflect the feedback while 15% agreed that they can reflect the feedback. For the question of whether the explanations for incorrect answers can be shared, 55% of the respondents answered that they cannot, whereas only 11% answered that they can. In addition, 51% of the students answered they received insufficient explanations from the system, whereas 14% said they received sufficient explanations. Finally, 48% of the students answered that the system providing explanations for incorrect answers did not enhance the learning effect, whereas 16% of the students thought it was helpful.

On the other hand, Table 1-B shows the survey results on the satisfaction level for the system in which a confidence scale and social feedback were provided. The response to the question on whether the explanation for incorrect answers reflected the level of the learners revealed that 65% of the respondents said the explanations reflected a level of learners, while only 7% did not agree. In addition, the response to the question on whether learners can find the reason for their incorrect answers through the provided explanation indicated that 64% said they found the reason while 8% answered no. Regarding the question on whether a learner can reflect the feedback about their incorrect answers, 72% of the students agreed that it is possible, whereas 7% did not agree. Furthermore, 66% of the students agreed that the system allows them to share their explanations of their incorrect answers, whereas only 2% did not. In addition, 60% of the students agreed that sufficient explanations can be provided through social feedback, whereas only 7% said they disagree. Finally, 76% of the students agreed that the system providing social feedback was
enhanced the learning effect, whereas only 4% did not agree. Based on the results in Table 1, the satisfaction level with regard to the learning effect for both systems can be found in Figure 5, in which the proposed SLMS had a higher overall satisfaction level of learning than SLMS without a confidence scale and social feedback with only explanations provided for incorrect answers.

Additionally, regarding the comparison survey question asking which system contributed more to their learning, 159 students indicated that the SLMS system which uses a confidence scale and social feedback was more helpful in increasing their learning, whereas only 39 students responded that the traditional LMS providing an explanation for incorrect answers only, but without a confidence scale and social feedback is more effective. An analysis of these 159 students’ survey results indicate five reasons why the SLMS system contributed more to a positive learning experience, as depicted in Figure 6. The most influential characteristic of the proposed system making a positive contribution to learning was providing various feedbacks through social network. In addition to the basic static explanations for their incorrect answers, students can obtain additional knowledge in relation to their learning because various opinions can be provided through an SNS. Similar to the first reason, the second most influential factor was about how to share information in relation to learning with close friends using an SNS. In addition, a real-time response, the use of a confidence scale, and active interaction with close friends were chosen as the remaining three advantages of the proposed system.

Figure 7 shows four representative factors influencing the learning effects of the system not providing confidence scale and social feedback. The biggest reason why 39 students selected this system was unreliable social feedback. Among them, 29 students pointed out that they cannot rely on the explanations provided through social feedback, which they saw as a drawback. On the other hand, the system without social feedback provided basic static explanations, which they viewed as reliable and satisfactory explanations that were more effective in increasing the learning effect. In addition, the students judged that the system without a confidence scale and social feedback was more effective because of its simple process and saving time.
As a result, more positive responses regarding the satisfaction level of the proposed system were received from the students when a confidence scale and social feedback where applied. Furthermore, the comparison survey also indicated that the proposed system had a higher learning effect than the other system. However, a small number of students who indicated that the system without a confidence scale and social feedback was better for the learning effect suggested that the quality and reliability of the explanations for incorrect answers provided through social feedback could not be guaranteed.

**Conclusion**

We proposed a SLMS based on SNS for providing learners with a variety of social feedback related to incorrect answers. In the SLMS, the accuracy of the learning results is increased by defining the confidence scale, and instead of providing a basic static explanation for incorrect answers to all learners, a variety of real-time social feedback is created and shared through the learner’s SNS friends based on Connectivism theory. For this purpose, we define the algorithm for processing the social feedback. For improving reliability of social feedback, the algorithm measures the intimacy between the learner and their SNS friends and the time of registered social feedback. Based on this, the algorithm ranks social feedbacks. Our experiment was to evaluate the SLMS performance in relation to collaborative learning through social feedback. The experiment participants were asked to study using the proposed SLMS method, both without and with a confidence scale and social feedback applied for two months. As a result, the evaluation result represents more positive responses regarding the satisfaction level from the experiment participants when a confidence scale and social feedback were applied. However, some participants were concerned regarding the reliability of the feedback.

As future work, we plan to extend the SLMS with added mechanisms regarding automatic filtering and management of unsuitable social feedbacks for improving reliability. Furthermore, we intend to apply the SLMS for other domains based on the positive responses received from the students.

**Acknowledgements**

This research was supported a Korea University Grant.

**References**


Group Formation in Mobile Computer Supported Collaborative Learning Contexts: A Systematic Literature Review

Sofiane Amara*, Joaquim Macedo², Fatima Bendella¹ and Alexandre Santos²

¹Department of Informatics, Faculty of Mathematics and Informatics, University of Sciences and Technology of Oran “Mohamed Boudiaf”, USTO-MB, Oran, Algeria // ²Centro Algoritmi, University of Minho, Braga, Portugal // sofiane.amara@univ-ustomo.dz // macedo@di.uminho.pt // fatima.bendella@univ-ustomo.dz // alex@di.uminho.pt

*Corresponding author

(Submitted February 13, 2015; Revised June 18, 2015; Accepted July 28, 2015)

ABSTRACT

Learners are becoming increasingly divers. They may have much personal, social, cultural, psychological, and cognitive diversity. Forming suitable learning groups represents, therefore, a hard and time-consuming task. In Mobile Computer Supported Collaborative Learning (MCSCCL) environments, this task is more difficult. Instructors need to consider many more issues, such as the rapid change of mobile learners’ context, their direct and natural interaction, and the characteristics of mobile devices and networks. This paper presents a systematic literature review (SLR) that examines the relevant solutions for the problem of group formation in MCSCCL environments. In the context of this SLR, an initial list of 178 papers was reviewed. After careful analysis of each paper using specific selection criteria and a quality assessment method, a final list of 12 relevant studies was filtered and used to answer the research questions. The findings revealed that: (a) there is a lack of approaches addressing the group formation problem in MCSCCL environments; (b) the most proposed solutions do not allow instructors to customize the grouping process; (c) there is no useful solutions to automatically capture and evaluate many of learners’ behaviours and context information; (d) the majority of approaches do not support a dynamic formation of learning groups; (e) the majority of approaches do not provide descriptions about the implemented grouping algorithms nor about the evaluation methods. Extracted and synthesized data from the selected studies is discussed in this paper, together with current research gaps and recommendations for further works.

Keywords
Learning group formation, Collaborative learning, M-learning, MCSCCL, Systematic literature review

Introduction

Collaborative Learning (CL) represents an essential educational approach defined by Dillenbourg as “a situation, in which two or more people learn or attempt to learn something together” (Dillenbourg, 1999). Henri and Lundgren-Cayrol (2001) consider CL as an active process by which the learner is working on the construction of knowledge, the instructor plays the role of facilitator of learning, and the group participates as a source of information, as a motivator, as a mean of self-help and mutual support, and as a preferred place of interaction for collective construction of knowledge. Many researchers demonstrated how CL is useful for improving the cognitive, psychological, and social development of learners (Dillenbourg, 1999; Zurita et al., 2005).

The development of information and communication technologies has led to the emergence of e-learning. It is a kind of learning based principally on the use of computers for constructing and delivering knowledge. Education researchers began then to search how to benefit from this technological evolution to improve the CL pedagogies. As result, since the late 1990s, a new branch of collaborative learning called Computer Supported Collaborative Learning (CSCL) has emerged.

Furthermore, the rapid development of wireless communication and mobile technologies enabled the emergence of a new form of learning termed M-learning. It allows learners through the use of mobile devices (PDAs, tablets, Smartphones, etc.) to learn anytime and anywhere, in formal or informal places. As result, CL has become possible in mobile situations and real world environments. Hence, Mobile Computer Supported Collaborative Learning (MCSCCL) represents a new paradigm of CL that is growing in use. This reality is confirmed by the big number of established MCSCCL projects (Yatani et al., 2004; Zurita et al., 2005; Boticki et al., 2011; Huang et al., 2014).

On the other hand, forming effective learning groups represents one of the important factors that determine the efficiency of CL. According to (Dillenbourg, 2002), studies show that three key conditions are required for any
successful CL: the task features, the communication media, and the group composition. The importance of learning group formation (LGF) process is confirmed by many researchers in the literature (Huang & Wu, 2011; Webb et al., 1998). Nevertheless, the social, cultural, psychological, and cognitive diversities of learners make the operation of forming suitable learning groups a hard and time-consuming task.

Although MCSCL represents a multidisciplinary research field (e.g., psychology, education, computer science), and although the importance of LGF process in succeeding the MCSCl activities, there is until date no effort to analyse the state of research on this topic. Thus, this paper reviews the studies addressing the topic of LGF in MCSCL using a systematic literature review (SLR) methodology. This SLR provides explicit information on what criteria are used for forming learning group, how to manually or automatically collect and use these criteria, how to support the dynamic formation of learning groups, what algorithms are used to form learning groups, and what methods are followed to evaluate the proposed approaches.

The paper is organised in four main parts. The first one shows how forming groups in MCSCL is compared to that of traditional environments. The second part describes the research methodology used in this work. Then, the main findings and gaps from this SLR, together with recommendations for further research are presented. Finally, our conclusions are provided.

Related research problem

One of the important questions raised in this research is whether the LGF approaches addressing the traditional environments are effective for MCSCL. To answer this question, it is important to know if MCSCL can be considered as only an extension of CSCL supported by mobile devices. The majority of researchers affirm that the answer is NO. M-learning context is different from that of more traditional e-learning (Parsons & Ryu, 2006). Additionally, Mobile collaborative applications do not replicate traditional learning scenarios, but they offer new learning opportunities, which cannot be reached without mobile technologies (Patten et al., 2006). Therefore, MCSCL does not mean “mobile + CSCL” (Looi et al., 2013); each paradigm has its particular environments, technologies, characteristics, practices and objectives.

MCSCL is highly dynamic in terms of users’ contexts. Mobile learners can dynamically obtain helps, recommendations, or learning content depending on their current context. For instance, when learners are near to a point of interest (POI), they receive information related to that POI, and when they move to another POI, the provided content is changed too. Contrarily, traditional environments are unable to adapt to dynamic changes of users’ context (current location, distances between learners, learning objects availability, etc.).

CSCL grouping algorithms are implemented to be run on high performances computers. Without Internet, those CSCL’s algorithms cannot be executed on mobile devices due to their weak technical characteristics. Therefore, new lightweight algorithms specific for mobile devices should be implemented and used with other types of networks such as Delay Tolerant Networks (DTNs).

CSCL’s learners are unable to communicate naturally, to move freely together, and to interact directly with learning objects. In the contrary, MCSCL’s learners are always in motion with face-to-face interactions. They may have different movement patterns (active, passive, etc.), levels of dialog and communication (social, shy, introvert, etc.), and preferences (preferred places, partners, learning objects, etc.). Those kinds of behaviourial information are not considered in traditional environments.

Unlike CSCL, MCSCL’s activities are generally exposed to many technical problems (disconnections, battery depletion, memory saturation, etc.); social problems (misunderstanding, disunion, selfishness, etc.); and natural/geographical problems (land degradation, weather changes, etc.). Those problems force the LGF approaches to be dynamic and able to (re-)form the groups in real-time. Such a mechanism of dynamically forming groups is completely ignored in traditional environments.

Taking into consideration those differences, one can affirm that using the same traditional solutions to form groups in MCSCL environments without considering their particularities could cause many problems such as, disunion of groups (e.g., when ignoring the learners’ geographical locations); demotivation, introversion, and isolation of
learners (e.g., when ignoring the personal traits of mobile learners, and their different learning and social behaviours); and obstruction of collaborative activities (e.g., when ignoring the different social, technical, and geographical problems that can happen in MCSCL environments).

Methodology

In this research, a guideline for performing systematic literature reviews (SLR) proposed by Kitchenham (2007) is followed. This methodology represents an efficient way to evaluate existing works relevant to particular research questions. The disadvantage of this method is that it necessitates more efforts than traditional methods of literature review. Figure 1 shows the followed steps to carry out this SLR.

![Figure 1. The followed steps for conducting the SLR](image)

Research questions

**RQ1: What are the learners’ personal characteristics used as grouping criteria?**

Learners’ personal characteristics could be used whether in mobile environment or not. But, the objective of using this RQ is to identify which personal characteristics are more used and more appropriate for MCSCL context, and how they are combined with the other grouping criteria (learners’ behaviours and contextual information).

**RQ2: Which learners’ behaviours are used as grouping criteria?**

Mobile learners may have different behaviours (they can be confident/afraid, active/passive, social/introvert, nervous/calm, etc.). This RQ aims to identify the different considered learners’ behaviours, and how the proposed MCSCL systems obtain, evaluate, and store this behavioural data.

**RQ3: Which kinds of context information are used as grouping criteria?**

Context is defined as any information that can be used to characterize the situation of an entity (e.g., learner, learning object, device). While MCSCL environments are known by their context awareness, the learning groups can be properly formed according to different context information. In this SLR, we search to identify which kinds of context
information are more considered, how this information is gathered from the mobile devices or other technologies (sensors, smart objects, etc.), and how it is used and combined with the other criteria to form learning groups.

**RQ4: How the dynamic grouping can be supported in MCSCL environments?**

MCSCL activities are generally exhibited to several problems that may lead to stop the collaboration between learners at any moment. For instance, the face-to-face interactions between learners may cause some social problems (e.g., selfishness, misunderstanding, disunion). Additionally, the technical limitations of mobile devices (connection rupture, battery depletion, memory saturation, etc.), and the geographical dispersion of learners in vast learning areas may break the communication between learners and destroy the learning groups. In such cases, the dynamic composition of groups plays an important role for quickly regrouping learners based on their updated information. The dynamic grouping processes can have two forms:

- Inter-sessions grouping: allows forming groups only before starting or after ending learning process. This kind of grouping is more useful for asynchronous mobile CL.
- Intra-session grouping: allows changing group members during the learning process. This kind of grouping is more useful for synchronous mobile CL.

Therefore, this RQ serves to identify the different solutions proposed for ensuring the two forms of dynamic formation of MCSCL groups.

**RQ5: What algorithms are used for forming learning groups?**

In MCSCL environments, the grouping algorithms do not only serve to use some stored learners’ data to form the groups, but it should capture and evaluate the learners’ behaviours, and interact with different components of the system to get instantaneous context information. Hence, this RQ aims to identify and examine the proposed or re-implemented LGF algorithms specific for MCSCL environments.

**RQ6: What methods are used to evaluate group formation processes?**

The best way to evaluate a MCSCL's LGF approach is to test it in real world context. However doing such experiments requires human resources (learners, instructors, etc.), material resources (mobile devices, networks, sensors, etc.), and a long time period. At the contrary, simulation methods allow researcher to assess their approaches several times with several settings and grouping criteria. However, modelling the different components of MCSCL systems (especially the human behaviours) represents another difficult challenge. Therefore, the answers of this RQ will show the different methods used to assess the proposed LGF solutions.

Beside those principal six RQs, this SLR extracts other information related to LGF problem, such as the types of learning groups, and the customization of the grouping processes.

**Literature sources and search terms**

The list of literature sources used in this SLR includes online databases, search engines, individual journals, and proceedings of scientific events (conferences and workshops) (Table 1).

<table>
<thead>
<tr>
<th>Resource type</th>
<th>Resource name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online databases</td>
<td>ACM, IEEE Explore, ScienceDirect, Scopus, SpringerLink, Web of Science,</td>
</tr>
<tr>
<td>Online search engines</td>
<td>CiteSeer, Google Scholar</td>
</tr>
<tr>
<td>Individual journals</td>
<td>Journal of Educational Technology &amp; Society, International Journal on E-learning,</td>
</tr>
<tr>
<td>Others</td>
<td>Proceedings of scientific events</td>
</tr>
</tbody>
</table>
After defining the list of literature sources, a search string was constructed using the following method (Kitchenham et al., 2007):

- Derive the major search terms (Table 2);
- Check the keywords list of already analysed papers to find more search terms;
- Identify alternative spellings and synonyms for each major term;
- Construct the search string using Boolean ORs to join alternative spellings and synonyms, and Boolean ANDs to join major search terms.

**Table 2. Search terms with their alternative spellings and synonyms**

<table>
<thead>
<tr>
<th>Terms</th>
<th>Alternative spellings and synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative learning</td>
<td>MCSCL, CSCL, Mobile learning; Ubiquitous learning; Pervasive learning.</td>
</tr>
<tr>
<td>Group</td>
<td>Team; Cluster; Set; Community; Assembly.</td>
</tr>
<tr>
<td>Formation</td>
<td>Construction; Composition; Organization; Constitution; Creation; Building; Assigning.</td>
</tr>
</tbody>
</table>

The major search terms used in this SLR are “Collaborative learning” and “Group” and “Formation.”

The resulting search string is as follows: (Collaborative learning OR MCSCL OR CSCL OR Mobile learning OR Ubiquitous learning OR Pervasive learning) AND (Group OR Team OR Cluster OR Set OR Community OR Assembly) AND (Formation OR Construction OR Composition OR Organization OR Constitution OR Creation OR Building OR Assigning).

**Studies selection**

To ensure identifying all relevant studies, this SLR used the following search method:

- Searching online engines and online databases;
- Searching manually scientific events proceedings and individual journals;
- Scanning the references lists of all found papers in order to avoid missing any interesting study.

By following this search method, an initial list of 178 papers was identified. However, this list includes some studies that do not address exactly the described research problem, or studies that are stored in multiple databases or published in many sources. Therefore, the objective of this stage is to eliminate duplicate papers and filter the relevant ones from the set of all found studies. The selection method is based on the use of the following inclusion and exclusion criteria.

**Inclusion criteria**

- If a study has both conference version and journal version, only journal version is included;
- If a study has many published versions, only the newest and the most complete version is included;
- If a study is stored in multiple sources, only one copy of this paper is included.

**Exclusion criteria**

- Papers that do not consider MCSCL environments;
- Papers that do not address the problem of LGF.

After applying those inclusion/exclusion criteria, a list of 20 papers was filtered (Table 3).

**Table 3. Distribution of found and selected studies in search sources**

<table>
<thead>
<tr>
<th>Source</th>
<th>Found studies</th>
<th>Selected studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>IEEE Explore</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>ScienceDirect</td>
<td>19</td>
<td>1</td>
</tr>
</tbody>
</table>
Scopus 55 3 
SpringerLink 12 1 
Web of Science 38 0 
Educational Technology & Society 5 4 
International Journal on E-learning 1 1 
Journal of Educational Data Mining 1 1 
International Journal of Learning Technology 1 0 
Other conferences and workshops 13 3 
Total 178 20 

Quality assessment

For well assessing and classifying the selected studies, a quality assessment checklist was developed (Table 4). Each paper was assessed independently by two authors. The quality assessment checklist is composed of nine questions labelled from QA1 to QA9, and each question is scored as follows:

QA1 is evaluated depending on the source of the paper:
- For conferences and workshops papers, the computer science conference ranking (CORE list) is used. The possible values are: A (1.5); B (1); C (0.5); No CORE ranking (0).
- For Journal articles, the Journal Citation Reports (JCR) is used. The possible values are: Q1 (2); Q2 (1.5); Q3 (1); Q4 (0.5); No JCR (0).

QA2 to QA9 should have one of the following values: Yes (1); Partially (0.5); No (0).

<table>
<thead>
<tr>
<th>ID</th>
<th>Question</th>
<th>Value</th>
</tr>
</thead>
</table>
| QA1 | Is the study published in a recognized journal or scientific event proceeding? | • CORE ranking  
• JCR ranking |
| QA2 | Is there a clear statement of the aim of research? | • Yes  
• No  
• Partially |
| QA3 | Does the study discuss any of related studies? | • Yes  
• No |
| QA4 | Does the proposed approach allow a dynamic grouping? | • Yes  
• No |
| QA5 | Does the study consider learners` learning behaviours? | • Yes  
• No |
| QA6 | Does the study consider context information? | • Yes  
• No |
| QA7 | Is the experimental procedure carefully explained? | • Yes  
• No  
• Partially |
| QA8 | Are the findings clearly stated and presented? | • Yes  
• No  
• Partially |
| QA9 | Was the paper cited by other researchers? | • Yes  
• No |

The quality assessment scores of each study given by two independent authors were saved and used to calculate the average score between them. Each study that has a quality score less than 5 was removed. At the end of this stage, a final list of 12 papers was obtained. The selected studies are labelled from S1 to S12 (Table 5). The quality assessment scores of each study are presented in Table 6.
Table 5. Selected studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Reference</th>
<th>Research questions addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Huang &amp; Wu, 2011</td>
<td>2 3 4 5</td>
</tr>
<tr>
<td>S2</td>
<td>Zurita et al., 2005</td>
<td>1</td>
</tr>
<tr>
<td>S3</td>
<td>Huang et al., 2010</td>
<td>2</td>
</tr>
<tr>
<td>S4</td>
<td>Messeguer et al., 2010</td>
<td>2 3 5</td>
</tr>
<tr>
<td>S5</td>
<td>El-Bishouty et al., 2010</td>
<td>1 3</td>
</tr>
<tr>
<td>S6</td>
<td>Hsieh et al., 2010</td>
<td>1 2 3 5</td>
</tr>
<tr>
<td>S7</td>
<td>Tan et al., 2010</td>
<td>1</td>
</tr>
<tr>
<td>S8</td>
<td>Giemza et al., 2013</td>
<td>1</td>
</tr>
<tr>
<td>S9</td>
<td>Mujkanovic et al., 2012</td>
<td>1 2 6</td>
</tr>
<tr>
<td>S10</td>
<td>Yin et al., 2012</td>
<td>2</td>
</tr>
<tr>
<td>S11</td>
<td>Yang et al., 2007</td>
<td>2</td>
</tr>
<tr>
<td>S12</td>
<td>Muehlenbrock, 2005</td>
<td>1 3 5</td>
</tr>
</tbody>
</table>

Table 6. Quality assessment scores

<table>
<thead>
<tr>
<th>Study</th>
<th>QA1</th>
<th>QA2</th>
<th>QA3</th>
<th>QA4</th>
<th>QA5</th>
<th>QA6</th>
<th>QA7</th>
<th>QA8</th>
<th>QA9</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>S2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>S3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>S4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>S5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>S6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>S7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>S8</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>S9</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>S10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>S11</td>
<td>2</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8.5</td>
</tr>
<tr>
<td>S12</td>
<td>1.5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Data extraction

To facilitate extracting the relevant data, each reviewer has used a data extraction form that lists the key information to be collected from each study (Table 7). Data extraction tables were then used to record the extracted data.

Table 7. Data extraction form

<table>
<thead>
<tr>
<th>Data ID</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>D01</td>
<td>Study identifier</td>
</tr>
<tr>
<td>D02</td>
<td>Name of author(s)</td>
</tr>
<tr>
<td>D03</td>
<td>Paper’s title</td>
</tr>
<tr>
<td>D04</td>
<td>Year of publication</td>
</tr>
<tr>
<td>D05</td>
<td>Paper’s type (Journal, conference/workshop proceedings)</td>
</tr>
<tr>
<td>D06</td>
<td>Quality assessment score</td>
</tr>
<tr>
<td>D07</td>
<td>Grouping type (homogeneous/heterogeneous)</td>
</tr>
<tr>
<td>D08</td>
<td>Whether the study supports a dynamic grouping or not</td>
</tr>
<tr>
<td>D09</td>
<td>Personal characteristics used as grouping criteria</td>
</tr>
<tr>
<td>D10</td>
<td>Learner’s behaviours used as grouping criteria</td>
</tr>
<tr>
<td>D11</td>
<td>Context information considered by the LGF process</td>
</tr>
<tr>
<td>D12</td>
<td>Used grouping algorithm(s)</td>
</tr>
<tr>
<td>D13</td>
<td>Methods used for evaluating the proposed solution</td>
</tr>
</tbody>
</table>
Data synthesis

Data synthesis phase serves to summarize and report the important results obtained from the analysis of the selected studies. To achieve this objective, the following strategy was followed.

- Answer individually the research questions by consulting the data extracted from the previous stage;
- Search additional findings besides the ones directly related to the research questions;
- Identify research gaps and provide recommendations for further research.

Results

Overview of studies

As shown in Figure 2, the majority of papers were published after 2005. This is probably due to the significant increase of mobile technologies’ usage in this period. For instance, the ratio of mobile cellular telephone subscription has reached from 20% in 2004 to 95% in 2014 (Figure 3) (World-Telecommunication, 2014). Although the mobile technologies are growing in use compared to the fixed ones, we cannot affirm that mobile CL is replacing CSCL. Each one of them has its particular characteristics, methods, practices, and objectives. For instance, we cannot develop complex applications using mobile devices, and we cannot perform location-based activities in real world context using desktop computers.

As shown in Table 8, the majority of studies do not combine the three types of grouping criteria in a single grouping process. By analysing the nature of learning groups in terms of homogeneity/heterogeneity of learners, this SLR classifies the selected studies into three groups: (a) Studies aim to form heterogeneous groups by maximizing the diversity within group (S1, S2, and S6); (b) Studies aim to form homogeneous groups by minimizing the diversity within group (S3, S8, S10 and S11); (c) Studies consider principally the criteria related to learning environments, and do not pay attention to the similarities/differences of learners (S4, S5, S7, S9 and S12). Generally, researchers in literature recommend the first type (heterogeneous grouping) to be beneficial, because it helps removing barriers between learners and improving their interaction and creativity (Hübscher, 2010). Other researchers find that it is useless to apply a specific grouping kind for all types of learners. Therefore, it is useful to leave the choice to instructors for selecting the nature of the groups according to different learning’s objectives, learners’ needs, activities’ types, etc.
Table 8. General features of the reviewed studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Grouping type</th>
<th>LGF criteria</th>
<th>LGF characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heterogeneous</td>
<td>Homogeneous</td>
<td>Personal characteristics</td>
</tr>
<tr>
<td>S1</td>
<td>✔️</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>S2</td>
<td>✔️</td>
<td>✗</td>
<td>✔️</td>
</tr>
<tr>
<td>S3</td>
<td>✗</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>S4</td>
<td>✗</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>S5</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>S6</td>
<td>✔️</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>S7</td>
<td>✗</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>S8</td>
<td>✗</td>
<td>✗</td>
<td>✔️</td>
</tr>
<tr>
<td>S9</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>S10</td>
<td>✗</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>S11</td>
<td>✗</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>S12</td>
<td>✗</td>
<td>✗</td>
<td>✔️</td>
</tr>
</tbody>
</table>

This SLR classifies also the reviewed studies according to their ability to customize the grouping process. We found that the majority of studies are not customizable. Instructors are generally unable to introduce their own choices for forming groups. For instance, they cannot select the nature of learning groups, the type and the number of grouping criteria, the number of learners in each group, etc. Instead, Studies S1, S2, and S7 provide some opportunities for customizing the grouping process. S1 allows choosing the number of groups and introducing a threshold of difference between learners. S2 enables (re)composing the learning groups by selecting the criteria that adapt the learning objectives. S7 allows learners to freely select one of two provided grouping options to be used by the algorithm.

Findings on the research questions

The following subsections present the extracted answers from the reviewed studies.

RQ 1: Learners’ characteristics used as grouping criteria

Since the majority of selected studies address MCSCL environments, the researchers pay more attention to criteria related to these environments and ignore the personal characteristics of learners. The most used learners’ personal characteristics are: gender (S6, S9), age (S7, S9), preferences (S2), interests (S3, S7, S8, S10), and learning backgrounds and experiences (e.g., languages, learning scores, learning capacities) (S5, S6, S7, S9, S11, S12).
RQ2: Learners’ behaviours used as grouping criteria

S1 proposes the use of an U-learning portfolio tool to collect and evaluate certain behaviours (see Table 9). The authors developed a systematic grouping scheme based on three successive processes: transformation of data from U-portfolio to a Portfolio Grid, building a learner similarity matrix, and using two clustering algorithms to distinguish learners into appropriate heterogeneous groups.

S3 presents a process of partners recommendation based on the analysis of learners’ reading interests. The proposed approach collects behavioural data from the social platform Del.icio.us, and creates a behaviour profile for each learner. Collaborative services use these profiles to recommend partners with similar interests and form appropriate learning communities.

S4 proposes a LGF approach based on a supervised machine-learning intelligent system. Depending on available contextual information (time, place, neighbourhood, etc.) captured from the learners’ devices, the system can automatically estimate the composition of learning groups.

S6 uses learners’ social interactions as the essential criterion to form learning groups. Highly interactive learners are identified as “hub’s learners,” and low interactive learners are identified as “island’s learners.” The two types of learners should be assigned evenly in each group.

S8 presents a mobile application for supporting the learners to form informal learning groups. The grouping mechanism uses learners’ profiles containing data about the course of studies, the number of semesters and the attended lectures. Based on this data the system provides recommendations to each learner.

S10 shows an approach for recommending partners according to the helping history between learners. The system could evaluate automatically the level of personal relationship between each pair of learners according to the frequency of the helping each other. Based on this information, recommendations of partners are provided to each learner.

RQ3: Context information used as grouping criteria

S1 proposes a location profile to record learners’ movements during their learning activities using Radio Frequency Identification (RFID) technology.

S4 considers some kinds of context information, such as: time and day of week, place, and list of neighbours (using Bluetooth’ MAC addresses). This information is combined with learners’ behaviours to train and test the machine-learning model.

S5 uses RFID technology to detect the surrounding objects. Information about detected objects in addition to the learners’ location allow the system to provide social knowledge awareness map for peer helpers.

S6 proposes a grouping solution based on the learners’ locations and their interactions. To detect the location of participants, each learner carries an ubi-coin (wireless detector), and each ubi-coin represents a node. If two nodes are within a specific distance for a moment, the system considers that they are in interaction.

S7 uses learners’ geographical locations as the principal criterion to form learning groups. The system uses certain technologies offered by the mobile devices (e.g., Wi-Fi, and Bluetooth, GPS) to obtain the learners’ locations.

S10 helps learners to construct a social learning network based on their location information collected through GPS sensors. This context information allows each learner to know the current location of other learners who have the similar interests. The system recommends then CL activities between them.

S12 presents an important solution allowing an ad-hoc creation of learning groups, using some context information (learner’s location, ambient sounds from PDA devices, and learner’s availability).
Table 9 summarizes the grouping criteria considered by the reviewed studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Personal characteristics</th>
<th>Learning behaviours</th>
<th>Context information</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Observing/Answering</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>Preferences/Academic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>Past activities (read</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>Preferences (places,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>Knowledge and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>Gender/Personal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td>Interaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S8</td>
<td>Course of studies/Number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td>Gender/Age/Motivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S10</td>
<td>Helping history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S11</td>
<td>Knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S12</td>
<td>Knowledge</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*RQ4: Proposed solutions to support dynamic group formation during the learning process*

S2 affirms that re-composing group members at any moment leads to reach many educational and social objectives. However, the proposed system enables only destroying the entire group and form new one with new members, and no solution was proposed to partially change group members.

The system proposed in S7 allows each learner to join or leave his/her group at any moment as his/her wishes. For instance, when a learner changes his/her geographical location, the grouping algorithm provides him/her dynamically with information about the existing groups in the proximity, and she/he decides to join or not each group.

*RQ5: Proposed LGF algorithm(s)*

S1 is the only study that clearly describes the used algorithms for grouping learners. These algorithms are classified into two types: heterogeneous grouping algorithm with given threshold of difference, and heterogeneous grouping algorithm with given number of groups. Based on the instructor’s choice, the chosen algorithm transmits data from an U-portfolio to a similarity matrix (M) in order to calculate distances between learners’ behaviours and then create groups based on these distances. We present in Figure 4 the successive steps used by one of the two proposed algorithms to form appropriate groups.
S4 uses data traces collected from the study of learners’ behaviours to train and test an intelligent system. The behavioural data should be transformed into input and output vectors to train IBL (Instance-based learning) and BayesNet (Bayesian Network) algorithms. These algorithms store the training data and use it later to predict the outputs at the arrival of a new contextual event.

S6 proposes an algorithm for recommending learning partners based on explored social networks. Through the analysis of collected data from learners’ activities, the interactive type of each learner (hub or island type) is identified. The proposed algorithm uses the learners’ interactive type in addition to some individual learning characteristics to assign learners to their appropriate groups.

S7 utilizes a grouping algorithm based principally on the geographical locations of learners. As grouping options, the algorithm could use two options: learning profiles together with learning styles of learners, or learners’ learning interests.

**RQ 6: Methods used for evaluating the learners grouping processes**

S1 used both simulation and experimental evaluation methods. Simulation was used to evaluate the proposed approach by comparing its average intra-cluster diversity (AID) with the AID of groups created randomly or according to academic achievement. A higher value of AID implies greater heterogeneity, and greater heterogeneity implies better results of heterogeneous grouping. Experimental evaluation method was used to evaluate and compare the learners’ behaviours. The learners are assigned to different groups using four grouping methods: random grouping; school achievement based grouping; grouping using the developed algorithm with given difference threshold; and grouping using the developed algorithm with given number of groups.

S2 uses an experimental evaluation method that compares pre-test and post-test results of a control group (formed using a random grouping) and three experimental groups (formed using three grouping criteria: preference, achievement, and sociability). The main objective of this experiment is to study the impact of different group reconfigurations on CL activities.
S4 proposes an evaluation method based on the use of same training datasets to create the learning model and to test the system. Second evaluation method uses only some parts of the training dataset to train the system, and other parts are used to test the system.

**Discussion and recommendations for further research**

Although mobile technologies facilitate gathering important information related to learning contexts (location, time, learners availability, learning objects availability, etc.), many considered studies (S2, S3, S8, S9, and S11) ignore completely the use of such context information. The learners’ geographical location is almost the only context information used as LGF criterion.

Through this SLR, we remark the lack of approaches that combine the three kinds of criteria (learners’ characteristics, learning behaviours, and context information) in a single process of LGF. Such a combination can make the grouping process more generic and adaptable to different learning contexts. Additionally, the grouping process should be customizable so that instructors can freely define the type, the number, and the weight of grouping criteria, together with different settings such as the number, the size and the type of learning groups. In such a way, instructor can customize the LGF process according to learning scenarios, type of learners, type of activities, needs, place, time, etc.

Many kinds of personal characteristics, learning behaviours, and context information have never been proposed as grouping criteria by the reviewed studies. Therefore, the following criteria are recommended to be considered in further approaches: Learner’s health status (healthy or disabled); Level of communication with instructors; Level of interaction with learning objects; Level of interaction with the system; List of disliked partners; Preferred times of learning; Availability of learners; Movement patterns of learners; Learning progression rate of learners; weather status; and learners’ agenda information.

Considering the greatest possible number of grouping criteria is a good solution to make the grouping process useful for any learning situation. However, selecting manually the proper grouping criteria could represent a difficult task. Therefore, developing new solutions that help instructors to know which criteria are appropriate to each situation is much requested. We propose, in this context, the use of machine learning algorithms that analyse the relationship between the past groups’ outcomes and the used grouping criteria, to form the best possible groups.

The intra-session dynamic grouping represents an essential requirement in MCSCL. However, only two papers (S2 and S7) propose some solutions for supporting the dynamic grouping. Such a dynamic grouping allows instructors to relocate at anytime the members of groups, or to form new groups depending on the task that is being performed in a given time and place. Additionally, it helps newly arrived learners, or learners who want to change their groups to quickly find new appropriate groups.

Another gap from this SLR is the lack of sharing the source codes of the implemented algorithms (except S1). This would prevent the MCSCL community to well compare, evaluate, and enhance the LGF tools and algorithms.

Through this SLR, we remark that many of reviewed studies did not evaluate their approaches. Some studies compared their automatic processes of LGF with manual or random methods (S1, S2). Some others used only questionnaires to evaluate the users’ satisfactions and obtain their feedbacks (S3, S5 and S8). Other studies used simulation methods to assess their approaches (S1, S4, and S9).

Among the gaps extracted through this SLR, is the lack of studies that consider the group leadership. Defining a leader for each group does not only help instructor for well controlling and communicating with the groups, but also facilitate the management of mobile CL activities within the groups. The group’s leader can define the role and location of each member; define the learning/working needs, objectives, and strategies; and find solutions to internal problems. Nevertheless, selecting the leader of each group represents another difficult task. This selection can be based on the analysis of the past learning outcomes and behaviours of learners.

Through this SLR, we remark that the majority of studies do not show how to obtain and measure the grouping criteria. We propose, therefore, a new mechanism for gathering, evaluating, and storing the values of grouping criteria (see Figure 5). Those criteria are classified into three groups:
• Learners’ personal traits, which could be introduced by the learners through their mobile interfaces, and stored in a specific database.
• Context information, which could be obtained in real time from the learners’ devices using some specific tools (e.g., GPS, RFID).
• Learners’ behaviours (Communication with partners, interaction with the system, etc.). To continuously evaluate those behaviours, we propose installing a set of event log files on the mobile device of each learner. Those files are considered as temporary databases that store different data related to the learners’ behaviours. The system can then examine those log files and extract the relevant behavioural information through a specific data extraction service. This extracted information should be stored in another active database (e.g., Oracle, MySql).

![Figure 5. Proposed mechanism for evaluating and storing the group formation criteria](image)

**Limitations**

The conducted SLR examines clearly the current state of research on the topic of LGF in MCSCL. However, some limitations have to be considered:
• This SLR addresses only the problem of LGF. Considering other related search problems, such as adaptive and personalised learning systems may provide other solutions for the LGF problem.
• This SLR focuses on the LGF problem in educational fields only. Considering other none-educational settings (e.g., formation of: business teams, sport teams, army troops) can provide more useful ideas for the topic of LGF.

**Conclusions**

This paper presents a Systematic Literature Review (SLR) on the problem of learning group formation (LGF) in Mobile Computer Supported Collaborative Learning (MCSCL) contexts. We believe that, through this SLR, we provided the MCSCL community a clear overview of the research on the LGF problem. In light of the extracted findings and research gaps, this SLR recommends the MCSCL community to: Develop new solutions for supporting intra-session dynamic grouping; Search new ways for making the grouping mechanism as useful as possible by considering the greatest possible number of grouping criteria; Propose new solutions to automatically select or recommend the relevant criteria that adapt to different learning aspects. Finally, and given the lack of considerable number of relevant approaches analysing the presented research problem, the MCSCL community are invited to pay more attention towards this issue by developing new contributions.

**References**


The Search for Pedagogical Dynamism – Design Patterns and the Unselfconscious Process

Agata Mouasher1* and Jason M. Lodge2

1Department of International Business and Asian Studies, Griffith University, Australia // 2Science of Learning Research Centre & Melbourne Centre for the Study of Higher Education, University of Melbourne, Australia // a.mouasher@griffith.edu.au // jason.lodge@unimelb.edu.au

*Corresponding author

(Submitted February 18, 2015; Revised July 7, 2015; Accepted September 2, 2015)

ABSTRACT

An apparent paradigm shift has created increased impetus to offer higher education across multiple delivery platforms. Utilising technology can support design and delivery for enhanced learning, albeit with additional pressures on academic workloads, affecting the ability to deliver quality formal education that meets the needs of individuals and society. The issue is exacerbated when technology, not pedagogy, drives decision-making, and further intensified by the formalisation of education. Using Mishra and Koehler’s TPACK framework, we argue that pedagogical dynamism is both necessary to maintain equilibrium of content-knowledge-pedagogy and a natural outcome. Further we suggest it is possible using Alexandrian design patterns and a return to the “unselfconscious process.” We critique existing design pattern work in education, and contribute a meta-theoretical exploration of Alexander’s principles and patterns to designing good-fitting forms impacting education. A scenario of designing for “online,” “on-campus” and “multi-mode” delivery of education is woven throughout to highlight implications for teaching practice.

Keywords
Pedagogical dynamism, Alexandrian design patterns, Educational technology, Educational design

Implications of technology-driven decision making in education design

Universities are embracing the growth of Internet access, implementing strategic plans to offer online learning (Kim & Bonk, 2006). In the U.S., nine percent of universities have MOOCs (Massive Open Online Courses) in various planning stages (Allen & Seaman, 2013). Technology has become a great ally in the task of designing education and enhancing learning. Students are now able to study via different, more flexible, delivery platforms. This trend is symptomatic of using educational technology in a bold move to pursue opportunities of converging, seamless teaching and learning. However, underpinning this trend, it is apparent that technology is driving decision-making about content and pedagogy. There are challenges of navigating through various technologies, trying to force-fit activities into online settings, or compromising activities due to technological constraints. What is increasingly missing is a focus on traditional course development and educational quality (Schmieder, 2008), or what Alexander (1979) termed “aliveness,” leading to “fragmented, alienating and/or dispiriting experiences” (Goodyear & Retalis, 2010).

Within a typical educational design context, the choice of what to teach is generally followed with technologies used and pedagogical goals (Mishra & Koehler, 2006). Take for example, the task of designing a program to be delivered both online and on-campus. University processes necessitate students choose a delivery mode at enrolment e.g., either “On-campus” or “Online.” As technologies are rarely designed with teaching and learning in mind, when such an operational/technological approach drives pedagogical strategy, then pedagogy and learning activities also become fixed by the chosen delivery mode (Goodyear, 2005; Means et al., 2010). Complexity arises in the design phase, such as the need to create different course profiles and pedagogies to suit each delivery platform (Figure 1). Also, the way students learn and teachers teach is not necessarily considered, making it more difficult to prioritise and react to individual student needs and offer a genuinely enhanced learning experience equivalent across the delivery modes. If a new mode for learning and teaching is then incorporated, it confronts basic design issues. A new delivery mode reconstructs the dynamic equilibrium between content, pedagogy and technology (Mishra & Koehler, 2006), creating additional pressure for educators to understand (Herring et al., 2014). Moreover, if the organisation adjusts processes and practices to meet the needs of the technology and not pedagogy, then everyone ends up serving “the machine” (Pollock & Cornford, 2005). The ability to make available a single set of materials and activities for...
all students, across all delivery platforms, accessible according to their circumstances, becomes increasingly complex (Phillips & Lowe, 2003; Flynn, Concannon & Ñi Bheacháin, 2005; Laurillard, 2009).

The following section unpacks the issue of a technology-driven approach amidst compounding, complex variables, using Mishra & Koehler’s (2006) TPACK framework, to illustrate the need to move towards a more dynamic pedagogical framework.

**The challenge of maintaining dynamic equilibrium as complex variables compound**

According to Mishra & Koehler’s (2006) TPACK framework, the three main aspects of learning environments - content, pedagogy and technology - must exist in a state of continuous dynamic equilibrium or tension, as in Figure 2. A change in Content Knowledge (CK), Pedagogical Knowledge (PK), or Technological Knowledge (TK), must be “compensated” by a change in the other two aspects to rebalance the state of equilibrium of good teaching.

Using our previous example of designing a course to be delivered “Online,” “On-campus” or “Multi-mode,” good pedagogy requires the Learning Objectives (LO) of a course be equivalent irrespective of delivery mode (Biggs & Tang, 2007; Meyers & Nulty, 2009), making the function CK (Learning Objectives) a fixed variable in TPACK. Similar reasoning is applied to TK as a fixed variable, because “Online” and “On-campus” learning respectively create the functions TK (Online) and TK (On-campus). If “Multi-mode” is also considered, the function TK (Multi-mode) must also be created. Furthermore, because each of these functions represents a specific delivery platform, each TK becomes a fixed variable.

This leaves Pedagogical Knowledge (PK). Any change in TK will therefore have to be compensated by a change in PK, because CK (LOs) is fixed (Mishra & Koehler, 2006, p.1030). For example, if TK (Online) has to change to TK (On-campus), then PK (Online) has to change to PK (On-campus). If a course is offered in all three delivery modes i.e., TK (Online), TK (On-campus) and TK (Multi-mode), the designer’s workload increases to design and maintain three or more PKs and their associated activities for each TK mode. For example, PK (Online) could include pre-recorded lectures available online, PK (On-campus) could be two-hour weekly lectures, and PK (Multi-mode) could offer a two-hour flipped classroom model. To complicate matters further, if a student’s individual needs are to be considered to achieve an enhanced learning experience, this adds an additional dynamic quality within the already dynamic equilibrium of the TPACK framework. For example, one student may choose TK (On-campus) and a few weeks into semester, change to TK (Online) because of choice or circumstance, whereas another student may choose to maximise both TK (Online) and TK (On-campus) for the course duration. Therefore, there is a need for PK to change dynamically to provide personalisation within each TK combination, thereby demonstrating the necessity of pedagogically-driven design with dynamic adaptivity to maintain equilibrium.
The TPACK framework suggests that technology is not separate from content or pedagogy. Educators and designers are needed with content area specialisation including the ability to model TPACK (Herring et al., 2014). The TPACK framework does not include scope for dynamically changing variables driven by student needs and contexts (Koehler et al., 2014), though other related models identify these e.g., the ICT-Related PCK (Angeli & Valanides, 2005). The ability to offer an equivalent experience for students while maintaining continuous equilibrium is difficult if not impossible. The delivery mediums are variable, as are the audience, student’s circumstances, pedagogical approaches, and technologies available (Herring et al., 2014; Means et al., 2010; Ury, 2004).

**Seeking dynamic equilibrium by returning to a pedagogically-driven focus**

Just like the flush toilet was once an innovative technology, in 100 years people will say the same about the Internet. Therefore, instead of pursuing educational design complexity by letting the technological environment, or TK, drive the design process, there is a need to return to a pedagogical focus (Mishra & Koehler, 2006; Goodyear, 2005; Garrison & Vaughan, 2008; Laurillard, 2009). Simultaneously, there is a push toward a learner-centric view, where educational technology is centered on quality of the learner’s activity, and design is able to capture experiences and reuse pedagogic forms (Goodyear & Retalis, 2010; Laurillard, 2009).

Ensuring educational quality is especially relevant when considering the need for dynamic educational design across different disciplines e.g., teaching medical procedures versus management principles, to overseas versus domestically enrolled students. One approach to maintaining continuous equilibrium with each change of any TK or PK variable (CK is fixed), while keeping a focus on pedagogy during both design and delivery, is to align it with a continuous quality improvement cycle e.g., PDCA or Six Sigma (Sokovic, Pavletic & Pipan, 2010). However, this approach is still static because changes may not always be possible as they arise, especially when organisational processes prohibit change (Pollock & Cornford, 2005). In the design phase, the TPACK model is a good tool to use, as it removes the focus on technology and provides a platform of equilibrium between student needs, pedagogy and technology. During delivery however, it is students’ individual needs that are most likely factor to unbalance the equilibrium. Therefore, in a learner-centric context, educators should be sensitive to student needs first, to which pedagogy responds, thereby influencing the technology to be used, as illustrated in Figure 3 as opposed to Figure 1.
In this ideal, dynamically adaptive educational delivery context, the complexity of maintaining equilibrium is intensified because the student’s needs drive changes in any of the forms of PK and TK. Mishra & Koehler (2006) and Goodyear’s (2005) models do not explicitly address knowledge of the actual teaching and learning process (whether the knowledge is held by the teacher or the student), nor knowledge of practical variables that will ultimately influence changes in TK, PK or even CK. However, this knowledge is just as important because “it is not only the result that is important, but the process [of learning] as well” (Alexander, 1964, p.133), and is wonderfully illustrated in Perry’s (1981) scheme of cognitive and ethical development. The need to create an environment that is conducive to pedagogical dynamism in order to maintain equilibrium prompts a search for other ways to effective design. A means to pedagogical dynamism can be found in design patterns and pattern languages.

**An exploration of design patterns for education**

The concept of design patterns originated in architecture in the 1970s, and has since been applied to disciplines involving complex tasks, such as software and aeronautical design. Educational design is also a complex task consisting of numerous layers (Meyers & Nulty, 2009; Schneckenberg, 2009; Laurillard, 2012; Mishra & Koehler, 2006; Palmer, 1998). Laurillard (2012) posits that teaching itself should be seen as a design science, inferring a solid foundation in using pattern languages in education. However, Alexander (1964) expressed Laurillard’s views more than 30 years earlier and this is seen in his own patterns on learning, teaching and education (Alexander, Ishikawa & Silverstein, 1977).

Numerous strides have been made in the field of educational design patterns (e.g., Mor, Mellar, Warburton, & Winders, 2014). However, more work is needed before such educational pattern languages can be used with confidence for pedagogical dynamism. There are three key limitations. Firstly, there is a scarcity in pattern literature specifically targeting pedagogy (Mor, Mellar, Warburton, & Winters, 2014; Mor, Warburton & Winters, 2012; Lyardet, Rossi, & Schwabe, 1998), pedagogical dynamism, or even the principles of the [educational] design process. Secondly, like any language, a pattern has vocabulary, syntax and grammar. Existing education design patterns vary immensely in how they are constructed and worded, and when viewed collectively, provide a measure of confusion for novice readers e.g., Gibbon’s (2010) design patterns are complex and difficult to understand without some knowledge of design layer theory. This defeats Alexander’s objective that “anyone can design and build any structure...” (Alexander, Ishikawa, & Silverstein, 1977). Alexander (1964; 1979) went to great lengths to identify how a pattern should be constructed, identifying a clear method by which patterns were rigorously tested over time. He also emphasised that “just like a flower cannot be made, but only generated from the seed” (Alexander, 1979, p.157), patterns are merely a means of connecting people to a deeper life-generating quality that he called “aliveness,” that can only be generated by the ordinary actions of people over time. There is little evidence to suggest that existing patterns in literature have been as rigorously tested and shown to link to that deeper life-generating quality Alexander alludes to. Thirdly, an in-depth study of Alexander’s work reveals that “not only the form of the results but the form of the path which led to them” is crucial (Alexander, 1964, p. 133). In other words, the
foundations underlying any factor that leads to the complexity of educational design, including our earlier functions of PK, CK or TK, are the processes of learning and teaching.

We realised it would be hypocritical to offer our own version of a pattern language until the knowledge of what makes up the process itself can be comprehensively understood. Therefore we sought a formula for pedagogical dynamism not in recent research, but in Alexander’s original and extensively tested work on pattern languages and design patterns.

**Alexandrian design principles for pedagogical dynamism**

According to Alexander, anyone can design and build any structure at any scale in a practical, safe and attractive way (Alexander, Ishikawa, & Silverstein, 1977), because there is a timeless way of building, if designers can rediscover the “unselfconscious process” (Alexander, 1964; 1979). Therein lies the fundamental underlying connection between patterns in a complex problem and the design of a form that could solve that problem. By stripping the complexity of the problem of design down to its basic elements, the process of design can be understood (Alexander, 1964). Even a designer who does not completely understand the problem at first can start the design process from any part of the problem which is understood, and from there work toward the unknown, producing a result that is usable.

Alexander (1979) identified a number of key underlying design principles i.e., understanding the principles of good form and good fit, in order to rediscover “aliveness” through the unselfconscious process, by being aware of the dangers of the self-conscious process. The following section explores each of these principles, progressing toward a deeper understanding of the basic tenets that underlie good design regardless of the level of complexity. From this, designers can then move onto design patterns as a template for delivering good-fitting forms that reflect a quality of “aliveness.”

**Understanding the principles of good form and good fit**

An ensemble is made up of various forms and contexts with varying degrees of fit, where context defines the problem in which form becomes the solution (Alexander, 1964). The ultimate object of the process of design within an ensemble is good form (Alexander, 1964). Yet one cannot design good form without understanding what constitutes good fit. For example, using the previous scenario of designing for “Online,” “On-Campus” and “Multi-mode” delivery, an educator’s ensemble may include the following forms: a weekly two-hour lecture, with a one-hour practical directly after; in a classroom setting to students enrolled “On-campus”; with live streaming technology to provide synchronous engagement with “Online” students; whilst recording the entire session for later asynchronous access by “Online” or “On-Campus” students, as part of the “Multi-mode” requirements. Form may be variable during design though not always during delivery. The context could be numerous organizational demands, such as the use of specific technologies, or restrictions of time and space (as per Figure 1). However, it can also be argued that arising student needs also create context, if a dynamically adaptive approach is considered (as per Figure 3). In fact, any externally driven variation in CK, PK, or TK becomes a boundary-changer to context. Within this ensemble, certain learning/teaching/assessment activities have better fit better under certain circumstances, such as exam revision having better fit at the end of a teaching period, or certain technologies fitting better with different knowledge domains, or live streaming technology having better fit for online engagement as opposed to face-to-face interactions.

Good form is therefore a designer’s ability to be alert for possible changes at every point in an ensemble, and simultaneously sensitive to the fit at boundary points within the ensemble, as boundaries of context can change (Alexander, 1964). Alexander argued that by shaping educational design within the demands that context places on it, designers can achieve “effortless contact or frictionless coexistence” (Alexander, 1964, p. 19), leading to a synthesised form dynamically as a solution to an existing complex problem i.e., a case for pedagogical dynamism. An example of this can occur in the above ensemble, where online students will log in to attend weekly live-streaming lectures, but not the practical sessions, thereby changing the boundary of context. However, because organizational processes do not allow form to be variable during delivery, no form changes can be made until the following semester, affecting the outcome of educational quality.
The Internet, the speed of change of technology, and the various student factors, constitute “impermanent materials and unsettled ways of life” which demand “constant reconstruction and repair” (Alexander, 1964, p. 50). Therefore, pedagogical dynamism is possible and necessary, not only as the outcome, but also as an element, of good form. However, there is no way of knowing, in education offered more flexibly which of the infinitely many relations between form and context to include or leave out, in order to match one good ensemble to another to create good form. This is because no matter how good the designer is, they are searching for a balance between two intangibles, one being a context that cannot be properly described, and the other being a form that has not yet been designed. However, because every designer is able to experience the sensation of “good fit” in an ensemble, good form can be achieved through good fit, which in turn can be identified through its negative, or “misfit” (Alexander, 1964).

Looking for misfits provides a way of identifying a solution within an ensemble where complexities exist. Educational design has long focused on listing requirements for good fit e.g., long lists of “principles of good practice,” rather than searching for misfits. If fit is treated as the absence of misfits, then a list of misfits become the criteria for good fit. Using the TPACK model, by identifying all examples of changing student factors, during design or prior delivery, designers can identify misfits e.g., not changing the PK for different TKs, instead making available the same set materials and activities for all modes of delivery, accessible regardless of student circumstances (Phillips & Lowe, 2003). Misfits can also be identified as they arise by creating a multi-layered ensemble of feedback loops enabling the assessment of PK, CK or TK at the point of a misfit, thereby maintaining equilibrium of good form.

Good teaching however, is more than just technique and understanding good form and fit. It requires the ability to design and teach as an expression of one’s own self (Palmer, 1998), or how Alexander (1964; 1979; 2004) describes living structures as “being” i.e., having spirit, in essence drawing linkages between environment, self and matter, and showed that good fitting form has a quality of “aliveness” or “wholeness,” which can be achieved through the “unselfconscious process.”

**Rediscover “aliveness” through the unselfconscious process**

Alexander categorised people into either unselfconscious or self-conscious cultures, based on methods of teaching and learning. An unselfconscious culture learns informally through continuous imitation and correction, with the unselfconscious process being self-adjusting and self-organising, hence providing active equilibrium with a system and consistently producing well-fitting forms (Alexander, 1964). In other words, in an unselfconscious process, the educational designer is also the one to deliver and be impacted by their own design as though they were the student. They are merely an agent, recognizing and responding continuously to misfits by making minor changes. Hattie (2009) alludes to this when he explains the need for teachers to become learners of their own teaching. Therefore, pedagogical dynamism is further possible when it is grounded in the conception that the designer, educator and student, together with teaching and learning processes, form an indispensable and mutually reciprocated relationship in a system that is unselfconscious. This approach creates a self-adjusting system that finds its own equilibrium.

Tacit knowledge is a good start, but as Bruner (1966) states, “instruction is a provisional state that has as its object to make the learner self-sufficient” (p. 53), and Perry (1981) further highlights the need for refined levels of skill and intuition to design teaching and curricula such that they invite, encourage, challenge and support students through the delicate yet inevitable nine stages of their ethical and cognitive development. Goodyear and Retalis (2010) reinforce this view, explaining that “good” educational technology design is characterised by a commitment to creating circumstances in which learning leads to experiences that are a source of pleasure, growth and transformation, as opposed to “fragmented, alienating and/or dispiriting experiences” (p. 18). Therefore, the unselfconscious process also recognises the need for a level of mastery [of a good teacher], because they intuitively know what works and what does not, and are able to react to individual student needs as they arise, changing pedagogy to suit. Technology is employed as an ally rather than being dictated by its constraints (Figure 3). However, time is required to develop this level of mastery (Ericsson, 1998). Furthermore, Alexander points out that “aliveness” cannot be attained by one person’s mastery alone, only when all the people in a community build something; with some master/expert at the helm; over time; sharing a common language rather than by “design at the drawing board” (Alexander, 1964; 1979).

Therefore, patterns for good form are to be found in the unselfconscious process of form-producing and thus responsible for its success. However, if the unselfconscious process contains the quality of “aliveness,” then
pedagogical dynamism will occur naturally due to the innate nature of the system to self-adjust and the inherent knowing of what needs to be done in the face of change. Therefore, pedagogical dynamism is the quality of aliveness, and if educational designers go back to the unselfconscious process, they might just unbundle some of this complexity and discover pedagogical dynamism. As Alexander admits though, society has lost the ability to recognise the unselfconscious process, due to having become self-conscious.

The dangers of a self-conscious process

Society’s need for individual expression and recognition of one’s individuality has brought about the self-conscious process, as each form created and its success is seen as individual achievement only (Alexander, 1964). What once took centuries of adaptation and development, a form clearly fitting its context, is beyond the average designer to achieve in a short time (Biggs & Tang, 2007; Fry, Ketteridge & Marshall, 2008). This is because the ever increasing number of factors that must fall simultaneously into place reduce the chances of success. In the unselfconscious process, one is directed without question to the right/wrong way. In the self-conscious process one is encouraged to question why, leading to the aggregation of failures/successes in some specific context. From this, good practice principles are generated, then theories and literary-based evidence, accruing ever more structure and abstraction. Formal teaching emerges, at which point the consequences of self-consciousness become visible on form. Goodyear and Retalis (2010) show that educational design is complex and challenging with often-unsatisfactory outcomes, highlighting that the self-conscious system’s lack of success lies not in individual lack of capability, but in actions following one’s awareness of one’s weakness to overcome this incapability (Alexander, 1964). In essence, by formalising education, the initial problem of designing good form has been complicated.

This is seen in formalised education, where organizational constraints govern effectively-dynamic design e.g., pre-set curriculums with sub-constraints within those contexts, such as a two-hour lecture here, or a 20-minute module there; the many variables needed to satisfy compliance with governments and accrediting bodies; and where young people are still being educated in “batches” and graduate with formal qualifications without the necessary experiential and practical proficiencies to succeed in the workforce (Robinson, 2010).

As a result of society having become self-conscious, a method is needed to convert a complex problem into good form. To do this, a conceptual framework needs to be invented, such as design patterns, to make explicit maps of the problem’s structure, which can lead designers to create form and thereby a solution (Alexander, 1964). Goodyear and Retalis (2010) also identify the shift to teaching-as-design, believing that educational design activities can be remarkably improved using a patterns-based approach. They too propose that good design “does not directly create activity, rather creating good learning tasks as blueprints for activity, leaving the detail of configuring the learning to the student” (Goodyear & Retalis, 2010, p.13). Good design, in particular, good design patterns, satisfy the self-conscious requirement for pedagogical dynamism.

Design patterns as a template for delivering good-fitting forms

A design pattern is a description of a recurring problem in an environment, that includes a description of the core of the solution to that problem, in such a way that the solution can be used many times over, without ever doing it the same way twice (Alexander, Ishikawa & Silverstein, 1977, p. x). Alexander showed that a good design pattern meets two empirical conditions: (1) the problem is real, meaning it can be expressed as a conflict among forces within the stated context, that cannot be resolved within that context; and (2) the configuration solves the problem, meaning that when the stated arrangement of parts is present in the stated context, the conflict can be resolved without side-effects (Alexander, 1979). Together, patterns form a language, with the power to generate an indefinite number of forms, with infinite variety in the details (Alexander, Ishikawa & Silverstein, 1977). Therefore, the question of the possibility of pedagogical dynamism can be restated into an Alexandrian problem: education becomes the stated context, including teaching and learning irrespective of the organisation. The system of forces can be those as outlined by Goodyear (2005) within educational design space, and in particular any changes in PK, CK or TK, such as between delivery mode, pedagogy, and individual student preferences. A likely conflict between these forces that recur is the prioritisation of delivery mode and resultant pedagogy over student needs.
Alexander developed a specific, detailed format for writing a pattern. Each pattern is consistently presented with other patterns, and also to “present the problem and solution of each such that it can be judged and modified without losing the essence central to it” (Alexander, C., Ishikawa, S. & Silverstein, M. 1977, p. xi). Alexander took a collaborative approach and many years designing and testing his patterns to “capture the invariant property common to all places which succeed in solving the problem” (p. xiv). Even then, Alexander admits to not having succeeded in this, and provides a rating for each pattern (using a series of asterisks) to denote the extent to which each pattern states a true invariant. Alexander’s pattern format is being used to some extent in educational design (e.g., Goodyear, 2005), but not across the board. Existing educational patterns from various emerging researchers, when read together, do not form a consistent view of educational design easily followed by a novice. Educators and designers should be critical of using or building on existing modern educational design patterns, without carefully investigating whether they have been created through a combined unselfconscious and self-conscious process, collaboratively, over time and thoroughly tested. Being aware of our own tendency toward the self-conscious, we highlight again that it would be hypocritical to blueprint our own design patterns for pedagogical dynamism. However, a pattern language to support pedagogical dynamism can still be created, using Alexander’s patterns. Extensive evidence of these patterns can be found in Alexander et al. (1973) and Alexander (1975) other publications. Figure 4 below lists nine patterns identified in Alexander’s work that impact education (lower numbers indicate higher-order patterns, where lower-order patterns are embedded in higher-order patterns). Together, these nine patterns form a dynamic pattern language for educational design, with an infinite number of final solutions having the quality of “aliveness.” If educators and designers follow the principles of the unselfconscious process, then any combination of the below patterns will, by design, produce a solution that is by nature pedagogically dynamic.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Network of learning *</td>
<td>To decentralise the process of learning and connect it with many people and places all over</td>
</tr>
<tr>
<td>43</td>
<td>University as a marketplace</td>
<td>To allow self-paced learning for all ages using the real work of professionals as the basic nodes in the network</td>
</tr>
<tr>
<td>80</td>
<td>Self governing workshops and offices **</td>
<td>To support the nature of humans and their work in giving people understanding through face-to-face contact and autonomous enough to let them govern their own affairs.</td>
</tr>
<tr>
<td>83</td>
<td>Master and apprentices *</td>
<td>To use the real work of professionals and tradesmen as the basic nodes in a network of learning</td>
</tr>
<tr>
<td>84</td>
<td>Teenage society</td>
<td>To encourage teenagers to work out a self-organised learning society of their own</td>
</tr>
<tr>
<td>85</td>
<td>Shopfront schools</td>
<td>To create a large number of work-oriented small learning centers in those parts of town dominated by work and commercial activity</td>
</tr>
<tr>
<td>148</td>
<td>Small Work Groups**</td>
<td>To create a workspace in such a way to make it possible for people to work in two’s or three’s, always in partial contact and partial privacy</td>
</tr>
<tr>
<td>151</td>
<td>Small Meeting Rooms*</td>
<td>To create the ideally proportioned distribution, by both size and position</td>
</tr>
<tr>
<td>157</td>
<td>Home Workshop</td>
<td>To encourage the formation of seminars and workshops in peoples’ homes</td>
</tr>
</tbody>
</table>

Figure 4. Alexander’s patterns that impact education

Alexander provides guidelines for designing a form-fitting solution. Using a chosen sequence of patterns in line with the unselfconscious process, the design process is both iterative and evolutionary. The designer starts working with a higher-order pattern, firstly envisioning how it might look; then letting the form grow gradually, from amorphous to more defined with each iteration of design; treating each pattern as an entire and whole entity, before moving onto the next. As each pattern begins to take shape, adjustments need to be made to other patterns, readjusting the total gestalt of the design, taking into consideration in higher-order patterns how lower-order patterns might be impacted, and in lower-order patterns that the integrity of higher-order patterns is kept. As much as possible, the design process should not be on paper, but directly on the project or site, as close to the live outcome, and as close to the end-user as possible (Alexander, Ishikawa & Silverstein, 1977).
Figure 5. Sample solution for pedagogical dynamism using Alexandrian patterns

Figure 5 illustrates how we approached a solution for pedagogical dynamism for the previously mentioned scenario of designing an “Online,” “On-campus” and “Mixed-mode” course, using Alexander’s design guidelines. We firstly chose five of the above patterns. Secondly, we envisioned what the pattern would look like and developed narratives describing each pattern’s key features. Thirdly, combining technological-content-pedagogical knowledge and mastery, over a series of collaborative peer-workshops, elements of good form began to emerge like a scaffolding (Meyers & Nulty, 2009) as we embedded into each pattern aspects of curriculum design such as learning objectives, teaching and learning activities, and assessment. The last diagram of Figure 5 shows a scaffolded visual summary of assessment activities aligning with learning objectives across the cognitive and knowledge dimensions of the learning process (Perry, 1981). Examples of boundaries of context that triggered continuous readjustments (during design and delivery) were where a learning activity touched the delivery mode, or where previous experiences of changing student needs touched any of the PK, CK or TK elements. Figure 5 outlines merely a portion of our design for a pedagogical dynamic course. Formative and summative feedback loops were established with students to test the effectiveness of the solution designed. The approach listed in this paper was implemented progressively over a number of teaching semesters.

However, as Alexander et al. (1977) states, “you cannot build a thing in isolation, but must also repair the world around it, and within it, so that the larger world becomes more coherent and more whole” (pp. xiii). The main limitations of such a pedagogically-driven approach were organisational processes, policies and structures preventing
higher-order pattern changes e.g., even though the Network of Learning pattern was envisioned to the full extent as per Alexander’s pattern, the eventual design was scaled down and embedded in just one course.

Conclusion

When education design is driven by technology, the result will be ill-fitting forms, which cannot maintain dynamic equilibrium of a system, and will perpetuate the issues currently seen in higher education and educational quality. By using the TPACK model, we have shown that dynamic equilibrium can only be maintained by returning to design driven by pedagogy, with pedagogical dynamism as both a necessary component and natural outcome.

Good-fitting forms are needed in design, beginning with a deep understanding of the process of learning and teaching, the process of design, and the principles of good form and good fit. These are fundamental tenets that generate the quality of “aliveness,” wherein patterns for good-fitting form occur naturally, because of the innate nature of the resultant system to self-adjust. An attuned designer will have the ability to be alert to, and dynamically respond to, possible changes (misfits) at every point in an ensemble. However, in the wake of compounding complex variables, the ability to design good-fitting forms is beyond the average designer partly because society has lost the ability to recognise the unselfconscious process, having unawares fallen victim to the self-conscious process. Therefore a method is needed to convert a complex problem, such as pedagogical dynamism, into good-fitting form.

So if progress is to be made towards education “aliveness,” tacit knowledge is a good start, but fundamentally, there must first be a conscious shift in introspective awareness, where designers and educators relinquish their identities and desires for recognition that define their work and self-conscious nature, and choose to rediscover the timeless way of teaching and learning. This may require changes to the individual and/or educational environments, such as: becoming learners of their own teaching; more sensitivity to identifying misfits and boundary-changes to context; more flexibility in changing teaching practices as needed, even if they don’t align with institutional demands; and more inclusivity of, and patience towards, students and their individual needs. Additionally, the aspect of mastery – a life of devotion to one’s craft – cannot be ignored nor fast-tracked, thus fervency in lobbying for organisational support of the unselfconscious process is also constructive.

Furthermore, in a meta-theoretical exploration of design patterns and principles, this paper provides a synthesis of existing literature about using design patterns to achieve pedagogical dynamism, because the field of design patterns specific to education is relatively new and still largely untested. So before adopting existing educational design patterns, designers should carefully assess them according to Alexander’s key guidelines, namely: the extent to which they have been created through a combined unselfconscious and self-conscious process; collaboratively; over time; sharing a common language rather than by “design at the drawing board”; written to a consistent structure and understandable to all; and thoroughly tested addressing Alexander’s two empirical design conditions. Clear instructions are offered in Alexander, Ishikawa and Silverstein (1977). In the meantime, educators and designers can still progress towards pedagogical dynamism by using Alexander’s original, well-tested patterns. We identify at least nine Alexandrian patterns directly impacting education, and provide a brief example of how this is possible with the scenario of designing a course for “Online,” “On-campus” and “Multi-mode” delivery using Alexander’s narrative approach.”

A key contribution of this paper is to highlight the need to strip the complexity of the problem of educational design down to its basic elements, and understand the process of design. Then even a novice designer can start the design process and produce a result that is usable. This paper transfers Alexander’s original design principles and patterns directly into the field of education, offering a synthesized, conceptual and authentic foundation for educational design. As there is a lack of literature on pedagogical dynamism, and what the “unselfconscious process,” as envisioned by Alexander, could possibly look like in modern education, further developments include: revisiting the integrity of existing education patterns in accordance with Alexander’s above-mentioned guidelines; increased collaboration or professional development workshops, to develop a shared common pattern language for education based on the above-listed Alexandrian patterns; and more literature of successful progress towards an unselfconscious process in education.

Any actions taken towards the unselfconscious process will achieve pedagogical dynamism as an inevitable by-product. Because educational design is not an isolated process, this will, in turn, create boundary-changes to a larger
context affecting farther-reaching issues, such as casualization of the academic workforce, increased academic loads, decreased focus on teaching as opposed to research, and academic disciplinary silos. The timeless way of teaching and learning goes beyond disciplinary boundaries, new technology, and even designing patterns and languages. They only serve as a reminder of what every teacher knows already when they give up their ideas and opinions, and do what emerges from themselves. Good-fitting forms, of which pedagogical dynamism is one, are not to be found in explicit rules, formalities, and institutionalization, but should, like life, unfold and evolve naturally, thereby continuously creating itself within a self-adjusting system. The process of teaching and learning can then become alive and joyful rather than complex and burdensome.

References


The Impact of Peer Review on Creative Self-efficacy and Learning Performance in Web 2.0 Learning Activities

Chen-Chung Liu1*, Kuan-Hsien Lu1, Leon Yufeng Wu2 and Chin-Chung Tsai3

1Graduate Institute of Network Learning Technology, National Central University, Taiwan // 2Science Education Center, National Taiwan Normal University, Taiwan // 3Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // ccliu@cl.ncu.edu.tw // maxlimitc@gmail.com // leonwu@ntnu.edu.tw // cctsai@mail.ntust.edu.tw

*Corresponding author

(Submitted February 23, 2015; Revised June 28, 2015; Accepted September 15, 2015)

ABSTRACT

Many studies have pointed out the significant contrast between the creative nature of Web 2.0 learning activities and the structured learning in school. This study proposes an approach to leveraging Web 2.0 learning activities and classroom teaching to help students develop both specific knowledge and creativity based on Csikszentmihalyi’s system model of creativity. The approach considers peer review as the core component in the Web 2.0 learning activities with the aim of engaging students in the creative learning paradigm. To gain a better understanding of the impact of such an approach on students’ confidence and performance, this study gathered and analyzed the works developed by 53 sixth graders in a Web 2.0 storytelling activity, as well as details of their creative self-efficacy. The results show that those students who experienced the peer review using a set of storytelling rubrics produced significantly more sophisticated stories than those who did not. Furthermore, the peer review did not exert a significant negative influence on the students’ creative self-efficacy. It was also found that the experimental group’s (students experiencing the peer review) creative self-efficacy consistently reflected their performance, while the control group’s creative self-efficacy did not. Such results support that the peer review process may help students to build a sophisticated level of reflection upon their creative work in Web 2.0 learning activities.

Keywords
Web 2.0 learning activity, Creative self-efficacy, System model of creativity, Field, Storytelling

Introduction

Creativity is considered as one of the core competences in contemporary education (U.S. Department of Education, 2010). Learning to be creative is thus infused into formal educational contexts to foster the creativity of students to address this requirement. It is considered important to develop effective practices and pedagogies for encouraging and enhancing students’ creativity in schools (Lassig, 2009). Recently, more and more Web 2.0 technologies, referred to as Internet-based applications, have been developed to empower users to interact and collaborate with each other as creators of user-generated content in an online community (O’Reilly, 2005). Web 2.0 technologies, such as blogs, wikis and social networking platforms, have been increasingly applied to promote open and creative learning experiences in various educational settings. For instance, the studies by Aragon, Poon, Monroy-Hernández and Aragon (2009) and Liu, Liu, Chen, Lin and Chen (2011b) have demonstrated the effects of such Web 2.0 platforms on augmenting creative activities, as they can afford a platform for students to create and share their creative works. Therefore, it is believed that the Web 2.0 technologies can enhance learner participation and creativity in educational settings (Greenhow, Robelia, & Hughes, 2009; McLoughlin & Lee, 2010; Ravenscroft, 2009).

However, many studies have pointed out the significant contrast between the creative nature of Web 2.0 learning activities and the structured learning that takes place in schools (Bennett, Bishop, Dalgarno, Waycott, & Kennedy, 2012; Mao, 2014). Formal education is restricted by the pre-defined curriculum in which students have to attain certain knowledge (Jimoyiannis, Tsiotakis, Roussinos, & Siorenta, 2013). However, the acquisition of the knowledge and assessment required in formal education may restrict the creative process. On the one hand, although the open and creative features of Web 2.0 learning activities may support active learning, previous studies have found that students’ work in such activities may be perfunctory and may lack critical construction of knowledge (Tess, 2013). On the other hand, the assessment of students’ performance, which is often conducted within a certain framework, may also be inconsistent with and interfere with the open and creative features of Web 2.0 learning activities.
It is widely believed that a critical pathway for developing students’ creativity is to model creative practices for students (Sternberg & Williams, 1996). This study thus attempted to propose an approach to leveraging Web 2.0 learning activities and classroom teaching to help students develop both specific knowledge and creativity. This approach is mainly based on Csikzentmihalyi’s system model of creativity (1999) which asserts that “creativity is a process that can be observed only at the intersection where individuals, domains, and field interact” (Csikzentmihalyi, 1999, p. 314). From Csikzentmihalyi’s perspective, individuals create new elements while knowing and operating old elements in a domain. However, not all elements are accepted as new elements; rather, new elements are “sanctioned by some group entitled to make decisions as to what should or should not be included in the domain” (Csikzentmihalyi, 1999, p. 315). On the one hand, these groups, i.e. the field, determine the value and originality of the new elements. On the other hand, individuals receive critical feedback to improve the elements they have created.

From the perspective of Csikzentmihalyi’s system model, the review process is a central component of the creative practice that students need to experience in order to understand the creative process. This study thus proposes integrating peer review activities with Web 2.0 learning activities to engage students in the creative learning paradigm and help them enhance their creative performance. In such peer review activities, students have the opportunity to reflect upon the creative learning process and to develop creative products through social interaction with others. Some empirical studies in the literature have pointed out the positive impact of peer review on creativity. For instance, Falchikov (1988) found that peer feedback helped promote student confidence and creativity in a project-based learning activity. Previous studies (Sluijsmans, Brand-Gruwel, & Merriënboer, 2002; Tsai & Liang, 2009) also found that students demonstrated an increasing level of creativity in their coursework through peer review.

Although the above studies showed positive impacts of peer review on student creativity, some other research has reported negative effects of Web 2.0 activities on the creative process. For instance, Windham (2007) found that creators may feel constrained about posting work online, and may think twice about the work itself and others’ perceptions of it because others can read it. Hurlburt’s study (2008) echoed this effect of Web 2.0 learning activities, indicating that the sharing of works is much more stressful than simple browsing and reading. This phenomenon is in part due to the fact that there is a profound link between peer commenting and psychological safety (Van Gennip, Segers, & Tillema, 2009), and the expectation of being assessed by others may influence one’s creativity (Smith, Michael, & Hocevar, 1990). There is thus an imperative need to understand how Web 2.0 learning activities and existing educational practices may be reshaped to fit each other in a way that can engage students in the Web 2.0 creative learning paradigm and assist them in critically constructing understanding of a discipline (Hemmi et al., 2009; Laru, Näykki, & Järvelä, 2012).

To gain a better understanding of peer review in the Web 2.0 context, this study attempted to investigate students’ creative performance and their confidence in a Web 2.0 learning activity. In recent years, researchers have begun to investigate the role of self-efficacy in engagement in the creative process. Creative self-efficacy, which shows individuals’ confidence in creating novel works, is particularly noted as addressing how it may influence individual and group creativity. It has been shown by several studies that creative self-efficacy significantly affects creative performance (Tierney & Farmer, 2002; Choi, 2004). However, creative self-efficacy has as yet received limited attention in the literature (Lassig, 2009). In particular, how peer review in the Web 2.0 learning activities may influence students’ creative self-efficacy in educational contexts remains unclear. The current study thus conducted an experiment to investigate the effects of peer review on creative self-efficacy and creative performance. By gathering the creative works developed by 53 elementary school students in a Web 2.0 storytelling learning activity, as well as an indication of their creative self-efficacy, this study aimed to investigate the following research questions:

- What are the impacts of peer review on students’ creative performance in the Web 2.0 learning activity?
- Does peer review impact students’ creative self-efficacy in the Web 2.0 learning activity?
- How is creative self-efficacy related to creative performance in the Web 2.0 learning activity?
Method

Participants

The participants of this study were Taiwanese elementary school students. Students at this level were chosen because creativity education is now receiving increasing attention in elementary schools. There is thus a need to examine how different pedagogies influence creative learning tasks. To evaluate how peer review influences students’ creative self-efficacy and performance in a Web 2.0 learning activity, two classes consisting of a total of 53 sixth graders from an elementary school in northern Taiwan were randomly selected. The participants did not receive specific training in storytelling skills as storytelling is not included in the elementary school curriculum in Taiwan. Furthermore, although creativity has been advocated by many educators, the school curriculum in Taiwan does not implement any formal course aiming to enhance students’ creativity directly. For example, the Chinese Literacy courses in Taiwan have been implemented to develop students’ Chinese literacy in vocabulary and reading without particular emphasis on storytelling. Therefore, this study investigated how the Web 2.0 platform could be implemented to facilitate elementary creativity education. Analyzing students’ activities on the Web 2.0 platform is helpful for understanding the effect of different Web 2.0 learning activities on elementary creativity education.

Procedure and the Web 2.0 storytelling activity

This study took a quasi-experimental approach to exploring the students’ creative performance and self-efficacy in a Web 2.0 storytelling activity. One class was assigned as the control group (n = 26; 13 males and 13 females) while the other was the experimental group (n = 27; 14 males and 13 females). Both groups used a Web 2.0 storytelling platform on iPads (described later) to create and share stories. They could also view the stories created by others on the platform. In addition, the experimental group students were provided with a set of rubrics (described later) to review the stories developed by others. Because the storytelling activity of this study involves narratives and drawings, the students’ scores of Chinese Literacy (i.e., abilities of Chinese literacy comprehension and expression) and Art from the previous semester were examined to confirm their basic skills related to the storytelling task. There was no significant difference between the two groups in their Chinese literacy scores (t = .067, p = .947) or art scores (t = 1.475, p = .146), indicating that the two groups did not differ in their basic storytelling skills in terms of the two required abilities prior to the experiment. Therefore, comparing the students’ creative self-efficacy and the creative performance of the two groups can help us to understand the influence of peer review in the Web 2.0 learning activity.

The experiment lasted for eight consecutive weeks and involved one hour per week of class time. During the first week, brief instructions on storytelling with the Web 2.0 storytelling platform were provided. The students were then asked to create a story entitled “Adventure to Mars” as an exercise during the 1st and 2nd weeks. The students in the experimental group were guided to evaluate other students’ creative performance during the exercise. They were then told that their stories created in the following weeks would be reviewed by their peers. This arrangement was to ensure that the experimental group fully understood how their stories were to be evaluated by their peers, and thus we can analyze how the peer assessment influenced the students’ creative performance. After the exercise stage, the students started a new story as their major project, “Saving the Forest,” from the 3rd to the 8th week. The actual review was conducted during the 7th and 8th weeks when most of the students had completed their draft stories. Rather than grading with a single score, the students were given a set of rubrics to review their peers’ stories. However, they did not receive any further training.

In order to facilitate the peer review of the students’ creative performance, a set of creation rubrics associated with storytelling was administered. Taking into consideration the time constraints and the children’s literacy capabilities, the set of rubrics given to the students was only a concise form of the storytelling performance rubrics developed by Liu, Chen, Shih, Huang and Liu (2011a) which were used by three raters (described later) in this study to evaluate the students’ creative performance in the storytelling activity. The concise form of the rubrics mainly evaluated the quality of the stories’ essential components such as the setting, events, actions and the ending, and further, considered a smooth and quality transition connecting across each of the components. In addition, the creativity level of these components was also evaluated. Lastly, the quality of technical performance such as drawings, voice narrations and framing were also included in the condensed rubrics. Hence, these rubrics, which translated the
original rubrics into five easy-to-understand items (see Table 1), were provided to the experimental group to help them review their peers’ stories in terms of both their content and technical quality.

<table>
<thead>
<tr>
<th>Item</th>
<th>Rubric</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The story has a vivid background, events, actions and ending.</td>
</tr>
<tr>
<td>2</td>
<td>The story proceeds smoothly in terms of its background, events, actions and ending.</td>
</tr>
<tr>
<td>3</td>
<td>The story is very creative in terms of its background, events, actions and ending.</td>
</tr>
<tr>
<td>4</td>
<td>The story is presented with very nice drawings, narrations, and sound.</td>
</tr>
<tr>
<td>5</td>
<td>The story is presented with very nice animation.</td>
</tr>
</tbody>
</table>

As a comparison, the control group did not experience this review process. However, they could freely view all the stories created by others and discuss the stories they were interested in in a face-to-face manner. All students were asked to complete a creative self-efficacy survey during the 9th week. The two groups’ feedback on the creative self-efficacy survey was analyzed to reveal the influence of the peer review.

The iPad storytelling platform

To facilitate the students’ participation in Csikzentmihalyi’s creativity process, in this study a Web 2.0 storytelling platform on iPads was developed (see Figure 1) to enable the students to create and tell a story. More specifically, the platform allowed the students to draw and use a voice recorder to tell stories. Furthermore, in order to support the review process in Csikzentmihalyi’s system model, the platform also supports Web 2.0 features. Notably, the Web 2.0 storytelling platform on iPads allowed the students to browse all the stories created by their peers using the sharing section (see Figure 2), as well as on a web-based platform. In addition to being able to browse all of the stories created by others, the students could also review the story created by a specific student. It was hoped that these features could construct a social platform wherein the students could experience the creativity process.

Figure 1. The interface of the iPad storytelling platform
To understand how the peer review influenced the students’ creative performance in the storytelling activity, all stories created by the students were evaluated by three reviewers. To reflect the students’ creative performance, the current study adopted the rubrics designed by Liu et al. (2011a) which provide a set of concrete guidelines to evaluate digital storytelling products based on Via’s rubrics (2002) which are commonly used for evaluating movie products. The rubrics examine the quality of stories with regard to nine dimensions: (1) transitions/edits, (2) planning/storyboarding, (3) action/dialog, (4) accuracy of information, (5) originality/creativity, (6) sound usage, (7) drawing, (8) camera picturing, and (9) framing. With 2-4 sub-items within each dimension, there are a total of 20 sub-items across the nine dimensions. An example sub-item within the dimension of transitions/edits is “This story proceeds smoothly in terms of presenting its background, events, actions, and the ending.” Another example within accuracy and dialog is “This story contains logical reasoning in its background, events, actions, and the ending.” The rubrics evaluated both content (dimensions 1-5) and technical (dimensions 6-9) quality so that we could obtain a better understanding of the stories developed by the students.

This study applied the consensual assessment technique (Amabile, 1982) to evaluate the students’ storytelling performance in order to assure the reliability of the storytelling evaluation. Three raters including one researcher of this study and two teachers in the elementary school used the aforementioned storytelling performance rubrics to evaluate all 53 stories. Pearson correlation analysis confirmed that the three raters’ grading yielded a highly positive correlation ranging from .80 to .93 (p < .01), indicating that the grading process was highly reliable. The grades from these nine rubrics were averaged together to indicate an overall performance of the students’ creative storytelling products. Additionally, the first five rubrics (i.e., transitions/edits, story planning/storyboarding, action/dialog, accuracy of information, and originality/creativity) were added together as the content quality score, and the last four (i.e., sound, drawing, camera, and framing) were added together as technical quality.

Furthermore, to understand the quality of the stories created by the two groups, we also examined how the students’ stories aligned with the important story elements (e.g., settings, events, actions and consequences) based on story grammar identified in the study by Liu et al. (2011a). Two coders examined each page of all stories together, and determined which of the four story elements each page should belong. However, a page might convey more than one...
element based on its content. When a combination of elements was seen and the coders were in agreement, the frame was regarded as having multiple story elements. The numbers of settings, events, actions and consequences which each story aligned with were obtained and analyzed to reveal the influence of peer review on story structures.

To identify how the peer review may have influenced the students’ creative performance, this study compared the two groups’ creative performance and their story structures with independent t-tests. Moreover, researchers have noticed the inadequacy of using only the result of statistical significance testing in statistical inference (Cohen, 1988). To address this issue, the Cohen’s effect size index \( d \) (1988) was also used to report the practical significance (effect magnitudes) along with each t-test. This is because the index can reveal the extent of the practical significant difference between groups for each variable in this study. According to Cohen’s rough characterization, \( d = 0.2 \) is deemed as a small effect size, \( d = 0.5 \) as a medium effect size, and \( d = 0.8 \) as a large effect size. Both the statistical significance obtained by the t-test and the practical significance were integrated to obtain more reliable results of how the students reacted to the two Web 2.0 learning activities.

**Analysis of creative self-efficacy survey**

The literature on creative self-efficacy indicates that an individual’s creative self-efficacy may not only relate to personal attributes such as one’s creative personality and the capability to generate creative ideas and products (Bandura, 1997; Amabile, 1997), but is also affected by socio-personal interactions (Hung, Huang, & Lin, 2008). In other words, creativity can either be encouraged or discouraged (Sternberg, 2012, p. 3), and is likely to be generated/eliminated by a socio-cognitive relational-self (Andersen & Chen, 2002). Therefore, the socio-personal interaction experience in managing significant feedback from their peers may influence the students’ confidence in their ability to perform creative tasks. With the purpose of investigating peer feedback effects on students’ creative self-efficacy, the current study adopted the survey developed in a series of studies by Hung et al. (2008) and Hung (2003). This survey includes 10 items thoroughly examining creative self-efficacy in three dimensions: “self-efficacy of the produced creative product,” “self-efficacy of creative thinking strategy,” and “self-efficacy of reactions to potential feedback.” The reported reliability of each dimension of the original survey (Cronbach’s alpha) was between .66 and .82. For the needs of the current study, a few items required modification to better align them with the condition of the storytelling experimental settings. The purposes and sample questions of the modified questionnaire are displayed in Table 2. The Cronbach’s alpha of the modified survey ranged from .64 ~ .80 among the three dimensions for another 144 students who also used the Web 2.0 storytelling platform for telling stories, indicating that the survey is adequately reliable.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Purpose</th>
<th>Sample questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creative products</td>
<td>To understand students’ perceptions of the product they created</td>
<td>I believe my own story is special when compared to others’. (Reverse)</td>
</tr>
<tr>
<td>Creative thinking strategy</td>
<td>To understand students’ perceptions of creative thinking and strategies</td>
<td>When a topic is assigned, I believe I can think of new ideas to create a good story. I can ingeniously apply some ideas or use sources to make creative outcomes.</td>
</tr>
<tr>
<td>Reactions to potential feedback</td>
<td>To understand students’ reactions to potential feedback given by peers</td>
<td>I will give up my ideas if my classmates or teacher criticize my story. (Reverse) I feel fine if others do not appreciate my story.</td>
</tr>
</tbody>
</table>

To obtain how the peer review may have influenced the students’ creative self-efficacy, the two groups’ creative self-efficacy was compared using independent t-tests and Cohen’s effect size. Moreover, to understand the interplay between creative self-efficacy and creative performance, the correlations between creative self-efficacy and storytelling performance for both the control and the experimental group were also analyzed. Such an analysis may further display whether creative self-efficacy might reflect their creative performance. It was hoped that the above analyses would reveal how peer review in the Web 2.0 learning activity can influence creative performance and creative self-efficacy.
Results

The influence of peer review on creative performance

Table 3 displays the comparison between the creative performances demonstrated by the students who reviewed others’ stories and those who did not. The results showed that the stories created by the two groups were significantly different in quality. The stories created by the experimental group received significantly higher scores than the control group for several rubrics. More specifically, their stories demonstrated a higher level of quality in their transitions/edits, story planning/boarding and accuracy of information. For the rubrics of originality/creativity and action/dialog, the experimental group also received marginally significantly higher scores than the control group. Because these rubrics were related to the content quality of the stories, these results indicate that those students who reviewed others’ work in the Web 2.0 learning activity were more likely to create quality content ($t = 2.87$, $p < .01$). The Cohen’s $d$ value for the overall content quality was .79 which almost reached a large effect size. However, the review process did not show a significant effect on technical quality. The Cohen’s $d$ value for the overall technical quality was .3, representing only a small effect size. Among the four technical rubrics, the experimental group students only received significantly higher scores than the control group in the drawing rubrics ($t = 2.32$, $p < .05$ Cohen’s $d = .63$). This may be because, with a clear storyline in mind, the experimental group could develop clearer drawings.

Table 3. The comparison of the Web 2.0 storytelling performance of the two groups

<table>
<thead>
<tr>
<th>Rubric</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Cohen's $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content quality</td>
<td>Experimental</td>
<td>27</td>
<td>3.06</td>
<td>0.86</td>
<td>2.87*</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.40</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitions and edits</td>
<td>Experimental</td>
<td>27</td>
<td>3.40</td>
<td>0.76</td>
<td>3.61**</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.64</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story planning and boarding</td>
<td>Experimental</td>
<td>27</td>
<td>3.13</td>
<td>0.83</td>
<td>3.26**</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.40</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action and dialog</td>
<td>Experimental</td>
<td>27</td>
<td>2.61</td>
<td>1.08</td>
<td>1.76</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.15</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy of information</td>
<td>Experimental</td>
<td>27</td>
<td>3.22</td>
<td>0.95</td>
<td>3.04**</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.42</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Originality and creativity</td>
<td>Experimental</td>
<td>27</td>
<td>2.86</td>
<td>0.97</td>
<td>1.83</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.40</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical quality</td>
<td>Experimental</td>
<td>27</td>
<td>2.58</td>
<td>0.80</td>
<td>1.00</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.37</td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sound</td>
<td>Experimental</td>
<td>27</td>
<td>2.98</td>
<td>1.21</td>
<td>0.86</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.72</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing</td>
<td>Experimental</td>
<td>27</td>
<td>3.22</td>
<td>0.68</td>
<td>2.32*</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.74</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camera</td>
<td>Experimental</td>
<td>27</td>
<td>2.09</td>
<td>1.04</td>
<td>-0.22</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>2.15</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Framing</td>
<td>Experimental</td>
<td>27</td>
<td>2.02</td>
<td>1.06</td>
<td>0.64</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>1.85</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

The structure of the stories created by the two groups was analyzed, as displayed in Table 4. The results show that the experimental group students created more versatile types of story elements than the control group did. They created an average of 6.26 story elements which nearly significantly exceeded the number of 5.23 for the control group ($t = 1.92$, $p = .06$). The Cohen’s $d$ was .53, representing a middle effect size. The analysis of the students’ stories also found that the stories created by the two groups showed some differences in their structure. While the control group created a relatively larger number of setting story elements, the experimental group created a larger number of event, action and consequence story elements. The Cohen’s $d$ for the Event was .47, which almost reached a middle effect size. Therefore, the above results support the argument that providing concrete rubrics for evaluating stories is helpful in terms of improving students’ content generation performance in Web 2.0 learning activities.
Table 4. The structure of the stories created by the two groups

<table>
<thead>
<tr>
<th>Story element</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Cohen d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>Experimental</td>
<td>27</td>
<td>1.41</td>
<td>0.89</td>
<td>-1.13</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>1.73</td>
<td>1.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event</td>
<td>Experimental</td>
<td>27</td>
<td>1.52</td>
<td>1.05</td>
<td>1.72</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>1.08</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>Experimental</td>
<td>27</td>
<td>2.22</td>
<td>1.50</td>
<td>1.59</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>1.54</td>
<td>1.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consequence</td>
<td>Experimental</td>
<td>27</td>
<td>1.11</td>
<td>0.64</td>
<td>1.28</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>0.88</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Experimental</td>
<td>27</td>
<td>6.26</td>
<td>1.63</td>
<td>1.92</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>5.23</td>
<td>2.23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The influence of peer review on creative self-efficacy

Although the provision of the review rubrics enhanced the students’ storytelling performance, it might have had a detrimental effect on their creative self-efficacy, as they realized that their stories would be reviewed by others. Table 5 displays the creative self-efficacy and the differences between the two groups. The results revealed that the two groups did not show significant difference in the three dimensions of creative self-efficacy. The Cohen’s d for the three dimensions ranged from .31 to .42, representing only a small effect size. The results reflect that both groups had perceived positive confidence in the storytelling experience in the three dimensions of creative self-efficacy. In particular, the two groups showed high confidence in the dimension of reactions to potential feedback (3.88 for the experimental group and 3.65 for the control group). In other words, even when they understood that their work would be evaluated by others, the students still perceived a high level of creative self-efficacy.

Table 5. A comparison of the two groups’ creative self-efficacy associated with the Web 2.0 storytelling activity

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creative product</td>
<td>Experiment</td>
<td>27</td>
<td>3.32</td>
<td>.67</td>
<td>1.56</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>3.04</td>
<td>.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creative thinking strategy</td>
<td>Experiment</td>
<td>27</td>
<td>3.56</td>
<td>.68</td>
<td>-1.47</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>26</td>
<td>3.27</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reactions to potential</td>
<td>Experiment</td>
<td>27</td>
<td>3.88</td>
<td>.85</td>
<td>-1.12</td>
<td>0.31</td>
</tr>
<tr>
<td>feedback</td>
<td>Control</td>
<td>26</td>
<td>3.65</td>
<td>.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The relation between creative self-efficacy and creative performance

To understand the interplay between creative self-efficacy and storytelling performance in the Web 2.0 learning activity, this study further analyzed the relationship between creative self-efficacy and creative performance associated with storytelling for the control and the experimental group. The results in Table 6 reveal that creative self-efficacy was positively correlated with storytelling performance in the experimental group. Examining the nine dimensions of the story quality, it was found that the experimental group’s creative self-efficacy is positively correlated with the storytelling performance in the dimension of originality/creativity ($r = .40, p < .05$) and drawing ($r = .47, p < .05$). By contrast, there were no significant relationships between the creative self-efficacy and storytelling performance in the control group. In other words, after experiencing the peer feedback activity, the students’ creative self-efficacy can more consistently reflect their actual creative performance than those students who did not experience the peer feedback activity. Such a result may reflect that students’ creative self-efficacy may be more related to their creative performance in the dimensions such as originality and creativity, revealing that the peer feedback with the rubrics may influence the students’ evaluation of creativity in a way that aligns with the teachers’ assessment, considering that there is a positive relationship between the self-reported creative self-efficacy and the creative performance evaluated by the teachers.
Table 6. The relation between creative self-efficacy and creative performance of the experimental and the control group

<table>
<thead>
<tr>
<th></th>
<th>Control group (N = 26)</th>
<th>Experimental group (N = 27)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>Content quality</td>
<td>-.07</td>
<td>.31</td>
</tr>
<tr>
<td>Transitions and edit</td>
<td>-.15</td>
<td>.32</td>
</tr>
<tr>
<td>Planning and storyboarding</td>
<td>-.11</td>
<td>.30</td>
</tr>
<tr>
<td>Action and dialog</td>
<td>-.02</td>
<td>.17</td>
</tr>
<tr>
<td>Accuracy of information</td>
<td>-.08</td>
<td>.32</td>
</tr>
<tr>
<td>Originality and creativity</td>
<td>.07</td>
<td>.40*</td>
</tr>
<tr>
<td>Technical quality</td>
<td>-.01</td>
<td>.29</td>
</tr>
<tr>
<td>Sound</td>
<td>-.01</td>
<td>.29</td>
</tr>
<tr>
<td>Drawing</td>
<td>-.15</td>
<td>.47*</td>
</tr>
<tr>
<td>Camera</td>
<td>.04</td>
<td>.12</td>
</tr>
<tr>
<td>Framing</td>
<td>.08</td>
<td>.14</td>
</tr>
</tbody>
</table>

Note. *p < .05.

Conclusions and implications

Many educators have emphasized the importance of integrating Web 2.0 technologies into classroom settings. Peer review may improve students’ understanding of a domain, but may also interfere with students’ creativity in Web 2.0 learning contexts. The results of this study found that peer review with an explicit set of storytelling rubrics enhanced the students’ storytelling domain knowledge as well as their storytelling performance. Moreover, it did not exert a significant negative influence on the students’ creative self-efficacy. Educators may find peer feedback helpful in terms of enhancing students’ creative processes in other Web 2.0 learning activities.

The results of this study may shed light on the consideration of incorporating Web 2.0 learning activities into classroom practice, particularly, how Csikzentmihalyi’s (1999) system model of creativity may be aligned with teaching practice in the classroom. One of the criticisms of the implementation of Web 2.0 learning activities in the classroom is the mismatch between the creative nature of Web 2.0 platforms and the highly structured nature of classroom teaching practice (Bennett et al., 2012; Mao, 2014). Students’ works on the Web 2.0 platform tend to be casual and lack critical knowledge construction (Tess, 2013). The implementation of Web 2.0 learning activities that fail to address this mismatch will not engage students in learning. A previous study by Liu et al. (2011a) also investigated students’ storytelling performance on a Web 2.0 platform, indicating that children tend to create only simple stories without developing sophisticated transitions. A follow-up study by Liu, Wu, Chen, Tsai and Lin (2014) indicated that providing hints to students improved their creative performance, but reduced their creative self-efficacy for participating in creative activities.

The current study applied Csikzentmihalyi’s (1999) system model of creativity as the theoretical framework to implement the Web 2.0 learning activities. The students were guided to play the role of reviewers to judge the creativity of their peers. It was found that this implementation influenced not only the content quality but also the students’ ability to structure stories in the Web 2.0 learning activity without hindering their creative self-efficacy. This result reveals that Csikzentmihalyi’s (1999) system model with clear quality rubrics to guide the peer review enhanced the students’ knowledge of storytelling, and such knowledge was further applied to develop their stories, such that the students in the experimental group exhibited better storytelling performance. These findings support the argument that the theoretical framework of Csikzentmihalyi’s (1999) system model of creativity is helpful for addressing the mismatch between teaching practice in classrooms and Web 2.0 creativity features, and for engaging students in learning.

This study also found that the implementation of Csikzentmihalyi’s (1999) system model of creativity in the classroom did not exert a significant negative influence on the students’ creative self-efficacy. It was found that the students who experienced peer review showed a similar level of creative self-efficacy to those who did not. These results contradict the findings of the studies by Windham (2007) and Hurlburt (2008) which indicated that social evaluation may cause anxiety that might restrict the creativity feature of Web 2.0 technologies. However, the findings
of this study should not be over-generalized to other Web 2.0 learning contexts as they were based on elementary school students’ feedback, while most of the studies regarding Web 2.0 learning activities such as those of Windham (2007) and Hurlburt (2008) concern higher education students’ experiences of using Web 2.0 platforms. Furthermore, it should be noted that the peer feedback rubric applied in this study was principle-based rather than rule-based, as it involved only general criteria such as creativeness and the smooth transition of the stories, and did not impose tight technical requirements such as the number of pages or the form of the stories. The study by Liu et al. (2014) indicated that providing detailed rules or hints during creativity processes may reduce students’ creative self-efficacy. Educators may need to take the form of review rubrics into account when implementing peer review in Web 2.0 learning activities. Future research is also required to examine how different forms of rubrics may impact creative self-efficacy and performance in Web 2.0 learning contexts.

In this study it was found that creative self-efficacy is correlated with story quality in the originality and creativity dimension. However, such a relationship was only found in the experimental group, not in the control group. These results partially align with the results of the study by Tierney and Farmer (2002), which indicated that students’ creative self-efficacy is significantly correlated with their creative performance. However, the control group students’ creative self-efficacy did not consistently reflect their creative performance. This may be partially due to the fact that the students in the control group did not gain sufficient evaluation skills associated with the storytelling task. By contrast, the experimental group’s creative self-efficacy consistently reflected their creative performance after they participated in the peer review activity. This finding may suggest that only when students have experience of engaging in social evaluation activities will their creative self-efficacy in the creative work be consistent with their creative performance. Educators may need to provide social evaluation opportunities for students during their initial encounters with creative work in the Web 2.0 learning context to build sufficient evaluation skills of creative work. For instance, teachers may organize an in-class discussion activity in which students may present their works and practice commenting on others’ work, or a structured peer review activity may be helpful to achieve this goal.

This study aimed to explore how peer review may affect creative performance and self-efficacy in a Web 2.0 learning context. It was found that the proposed peer feedback may enhance the storytelling performance without hindering the creative self-efficacy associated with storytelling. However, the study was conducted in a real classroom where the students knew each other. It is not clear how peer review might influence the creative performance and self-efficacy in purely online contexts where participants do not know each other. Furthermore, it should be noted that this study was only a small-scale investigation and mainly focused on storytelling tasks. The subjects of this study were students in an elementary school in Taiwan. Cultural differences may have a potential impact on students’ creative performance and self-efficacy. It would be interesting to see how Web 2.0 learning activities could influence the creative process of students of different ages and countries. Nevertheless, the peer review mechanism is common in many online Web 2.0 learning activities. Thus, having acknowledged the limitations of the sample size and the creative task, this study can still provide potential approaches to using Web 2.0 technologies to enhance learning activities in the settings of other domains.

Acknowledgements

This research was partially funded by the Ministry of Science and Technology, R.O.C. under 103-2511-S-008 -014 -MY3 and 101-2511-S-008 -005 -MY3.

References


Academic Achievements and Satisfaction of the Clicker-Aided Flipped Business English Writing Class

Yu Zhonggen1, 2* and Wang Guifang2
1School of Foreign Languages, Hohai University, Nanjing, China // 2English school of Zhejiang Yuexiu University of Foreign Languages, Shaoxing, China // 18951801880@189.cn // 406302260@qq.com
*Corresponding author

(Submitted March 18, 2015; Revised July 27, 2015; Accepted August 21, 2015)

ABSTRACT

The flipped classroom has been achieving a great success in teaching innovation. This study, aiming to determine the effectiveness of the flipped model in business English writing course, combined the quantitative with the qualitative research methods. Participants were randomly selected from undergraduate students majoring in business English. The research instruments in this study included a satisfaction scale, a Business English Writing Test, and a semi-structured interview. The research procedure was made up of a pretest—treatment—posttest design. Both hypotheses were accepted and it was concluded that (1) the flipped business English writing classroom brought about better academic achievements than the traditional one, and (2) the flipped business English writing classroom was more satisfactory than the traditional one. Future studies could pivot on different courses and expand the research scopes to examine the effectiveness of the flipped classroom.

Keywords

Clicker-aided flipped classroom, Traditional classroom, Academic achievements, Satisfaction, Business English writing

Introduction

The new century is witnessing a substantial change in pedagogical innovation, among which the flipped classroom has been catching public attention. The flipped classroom flipped the teaching process where the delivered contents conducted in traditional class were moved outside class. In-class activities which needed the peer collaboration guided by the teacher in the flipped class were conducted before and after the traditional class (Ojalo & Doyne, 2011). The flipped classrooms required students to learn and internalize the knowledge by watching videos and reading other related lecturing materials provided by the teacher before they physically entered the class to share difficulties, questions, and other unsolved problems with peers and the teacher. The teacher also encouraged learners to participate in in-class activities such as group discussion, academic presentation and gaming so that the class functioned as a platform for peer discussion and problem solving (Triantafyllou & Timenko, 2014). Most of the studies on flipped classrooms could not achieve success without the aid of information technologies. This study, aiming to study flipped classrooms aided with clicker system—a frequently used information technology, is therefore meaningful and progressive.

Literature review

Satisfaction of the flipped classroom

A great number of studies focus on the satisfaction of information technology aided flipped classrooms (ITFC). ITFC could improve learners’ satisfaction and enhance their self-efficacy; hence they tended to hold positive attitudes towards the flipped classroom. The flipped classroom could satisfy learners because it was proved effective in engaging students, and could improve their self-efficacy in independent knowledge learning (Enfield, 2013). The flipped model was effective in learning and thus satisfactory to students. Under the flipped model, students’ attendance improved with more efforts paid (Chen et al., 2014). Through a survey and some interviews, it was found that most of the students positively assessed the flipped classroom because of its easiness to pause, rewind, and review lectures, together with encouraged self-directed learning and improved instructor availability (Schultz et al., 2014).
Moreover, students who enrolled in information literacy competency and critical thinking also reported that not only students but also teachers positively evaluated the effectiveness of the flipped classroom, which was identified through semi-structured interviews (Kong, 2014). The flipped classroom also enhanced learner empowerment, development, and engagement (McLaughlin et al., 2013). And students thought that the self-learning before class obviously improved in-class learning activities. Many students such as those enrolling in undergraduate nutrition courses preferred the flipped classroom to the traditional one (Gilboy, Heinerichs, & Pazzaglia, 2015).

Flipped classrooms could strongly motivate learners to join learning activities (Tune et al., 2013), so that learners tended to be satisfied with the flipped method (Pierce & Fox, 2013). Students who enrolled in statistics course were also satisfied with the flipped teaching model (Wilson, 2013), and students and faculty both believed the flipped model was satisfactory (Gilboy, Heinerichs, & Pazzaglia, 2015). Combining flipped pedagogic model with online NextGenU courses was also satisfactory in terms of public health higher education, where despite that students obtained the same learning outcomes, the flipped classroom received more satisfaction than the traditional one (Galway et al., 2014).

**Academic achievements of the flipped classroom**

The recent decade has witnessed many studies on the flipped classroom (Moffett & Mill, 2014), where the majority advocated the flipped classroom and thus the flipped innovations has been widely accepted especially in primary and secondary education (Slomanson, 2014). A well designed flipped model could be helpful for students to obtain favorable academic achievements (Triantafyllou & Timcenko, 2014). Students could perform significantly better in the flipped classroom than those through the traditional teaching approach (Schultz et al., 2014). The flipped pedagogic approach was a great success for both students and faculty (Critz & Knight, 2013).

Students joining the flipped classroom could obtain better academic achievements than those in the traditional classroom. Students under flipped models obtained a significant growth in information literacy competency and critical thinking abilities (Kong, 2014), a quantitative and analytical course (Asef-Vaziri, 2015), retention rates and the self-learning package of radio-graph (Leung, Kumta, Jin, & Yung, 2014), physiological knowledge learning among graduates (Tune et al., 2013), the courses of pharmacy (McLaughlin et al., 2013) and statistics (Wilson, 2013) through reasonably designed flipped models, active dialogues, and learner autonomy.

The flipped model could increase students’ class attendance rate, encourage students’ participation in learning activities, enhance learner self-efficacy, improve learning outcomes, and thus prepare students for 21st-century healthcare needs (McLaughlin et al., 2014). In addition, the flipped model was especially effective in relatively isolated places due to its flexibility of learning venues (Butgereit & Osman, 2014).

The flipped model encouraged students to learn independently, stimulating problem-based, discovering and individualized learning strategies (Sengel, 2014). The online engagement and homework rates were increased among at-risk students who received flipped teaching, through which students increased their online engagement, homework completeness rates and success in course learning. Students’ class failures in renal pharmacotherapy (Pierce & Fox, 2013), mathematics, English, science, and social studies were eliminated (Plumerfelt & Green, 2013) as well.

In-class learning activities such as peer discussion, group discussion and gaming promoted students to interact in class with the teacher and themselves. The flipped class could be activated and thus dynamic by using educational technologies such as i>clickers (Thoms, 2012), multimedia projection, and other digital learning platforms, increasing academic achievements (Tune et al., 2013). Furthermore, the initial contact with flipped classrooms and learning results were positively correlated. In other words, the earlier students received flipped pedagogic approach, the better academic achievements they would obtain and the easier it would be for them to accept projects or tasks needing self-initiated learning (Leung, Kumta, Jin, & Yung, 2014). To sum up, the flipped classroom could lead to significantly better academic achievements than the traditional teaching approach.

Although a huge amount of literature was conducted on the flipped classroom, very few of them focused on the effectiveness of the flipped business English writing classroom. This study aims to identify whether the flipped classroom outweighs the traditional one in terms of academic achievements and satisfaction. Research questions can be divided into two generic ones, either of which involves a few specific sub-questions.
Will the flipped business English writing classroom bring about better academic achievements than the traditional one? This generic question is supported by five specific questions, i.e., (A) can the participants more satisfactorily finish the writing task based on all the requirements through the flipped pedagogy than the traditional? (B) is language used more formal and natural through the flipped pedagogy than the traditional? (C) are there significantly less language errors through the flipped pedagogy than the traditional? (D) is the language structure more various and the vocabulary range larger through the flipped pedagogy than the traditional? (E) are the register and format of writing more proper through the flipped pedagogy than the traditional?

Will the flipped business English writing classroom be more satisfactory than the traditional one? This generic question includes 14 specific questions, e.g., (A) are the participants able to access a computer with an Internet connection to do their work for the flipped class? (B) are the participants satisfied with the degree of contact they have with their teacher when working through the flipped class? (C) are the participants pleased with the success they are having with completing the flipped class? (D) is the participants’ technology knowledge level sufficient for learning in a flipped environment? (E) are the participants feeling somewhat connected to the University setting by taking a class that places emphasis on learning through the flipped classroom? (F) would the participants prefer to take more of their classes through Internet delivery? (G) Do the participants believe that working in a flipped environment enables them to play a more active role in the learning process than in a traditional one?

Based on the above research questions, research hypotheses are raised as: (1) The flipped business English writing classroom will bring about better academic achievements than the traditional one. (2) The flipped business English writing classroom will be more satisfactory than the traditional one.

Research methods

In case both quantitative and qualitative research methods are designed to examine a particular subject area of interest, there will be a great possibility of neutralizing the flaws of one method and strengthening the benefits of the other for the better research results although both of them have strengths and weaknesses (Hussein, 2009). Many scholars considered it important to combine both research methods and even some claimed it to be a third research method (e.g., Bryman, 2004; Creswell, 2003; Tashakkori & Teddlie, 2003). Therefore, this study combined quantitative with qualitative research methods in order to maximize the reliability of the results.

Participants

Participants were all business English majors with the same business English writing skills, which was confirmed by daily performance, assignment, mid-term and final exams, plus a pretest prior to the experimental teaching. The sample, 69 undergraduate students, was randomly selected to stand for the target population who learned the course “Business English writing.” They were between 19 and 23 years old (mean age = 21.33; SD = 1.172). Among them, 34 were males, and 35 were females. They were divided into two parts and placed in Class A and Class B respectively. All the participants reported that they were able to deal with online learning technologies used in this study.

Research instruments

The research instruments in this study included a satisfaction scale, a Business English Writing Test, and a semi-structured interview.

The satisfaction scale

The satisfaction scale adopted in this study included 14 questions (see Appendix A), which were mainly adapted from Stokes’ scale (2001). The scale was proved both internally consistent and valid to identify the satisfaction level.
(Stokes, 2001). Each question was followed by a five-Likert scale, ranging from “never” to “very often” and contributing to one to five points.

**The Business English writing test**

The Business English exam (or Business English Certificate), shortened to BEC, refers to the Cambridge Business English exam. BEC provides learners with Cambridge Business English Certificate, examining English communication skills in the real working environment. It is widely accepted by many global enterprises and educational institutions as the entrance certificate of English language proficiency. BEC was introduced by the Examination Center of Chinese Ministry of Education in 1993. After years of practice and promotion, BEC, having been well-known in China, is presently a powerful language proficiency certificate in the business career.

BEC test is classified into elementary, intermediate and advanced levels. Participants in this study received writing tests on the intermediate level and received BEC intermediate teaching based on their actual English proficiency. BEC writing tests include two parts. The 1st part required participants to write an essay within 30-40 words. Assume that the participant works for a company which is going to buy a set of equipment from China, and he or she is required to translate a lot of specifications and instructions within four months, which is a harsh job. Therefore, he or she decides to advertise for two experienced translators as soon as possible. He or she is required to write a short note to Mr. Max Remington, the Public Relation’s manager and asks for an advertisement for two translators, where the reason and the urgency should be stated.

**Table 1. Scoring mechanism for BEC**

<table>
<thead>
<tr>
<th>Level</th>
<th>Writing details</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>All the writing requirements are met.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The language is formal and native with few errors.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>The structure is of varieties and the range of vocabulary is large.</td>
<td>18-20</td>
</tr>
<tr>
<td>The sentence is properly organized.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The register and format are proper.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All the writing requirements are nearly met.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The language is largely formal and native with a few minor errors.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>The structure is reasonable and the use of vocabulary is correct.</td>
<td>14-16</td>
</tr>
<tr>
<td>The sentence is largely properly organized except some logical problems.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The register and format are largely proper.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All the writing requirements are met except some minor contents.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A few errors exist but readers can read through.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>The structure and vocabulary are both moderately commanded.</td>
<td>10-12</td>
</tr>
<tr>
<td>The organization is generally organized.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The register and format are OK although not so successfully used.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All the writing requirements are not fully met. Some minor contents are missing.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some errors influence communication and readers feel confused by too many errors.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>The structure and range of vocabulary is limited.</td>
<td>6-8</td>
</tr>
<tr>
<td>The sentence is not well organized, leading to misunderstandings.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The register and format are not proper.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major contents are missing.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The language is seriously misused.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>The structure range of vocabulary cannot meet the requirements.</td>
<td>2-4</td>
</tr>
<tr>
<td>The sentence is too poorly organized to understand.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost no proper register and format are used.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No writing requirements are met.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>The number of words are fewer than 25% of the required.</td>
<td>0</td>
</tr>
<tr>
<td>Completely unperceivable.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completely irrelevant.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part 2 requires the participant to read the chart which shows the trend of consumer confidence, real hourly wages, employment and credit development, plus given years. The other numbers, clearly shown by the lines in the chart,
are not directly presented. Participants are required to use the information in the charts to write a report (about 100-120 words) analyzing the reason why the consumer confidence increases.

After one semester’s learning, participants in this study were required to complete both parts scored strictly based on the scoring criteria stipulated by the exam committee. The committee scored the writing based on (1) scope of the topic, (2) formality and smoothness of the language, (3) language correctness, (4) the language structure and vocabulary range, and (5) the register and format of writing. The full score was 20 points, which was divided into six levels ranging from 0 to 5. The detailed scoring mechanism is shown in Table 1.

The semi-structured interview

In order to minimize the drawbacks of the quantitative research method and maximize the advantages of the qualitative one, a qualitative research was also conducted through an interview. To fit and confirm the quantitative research, the questions of the interview mainly focused on the research hypotheses put forward in the quantitative research process, where both satisfaction and academic achievements of the flipped classroom were highlighted, coupled with some related issues to relax the interviewees. Examples are “How is everything with you here?” “What’s the weather like today?” and “What do you think of the environment here?” Questions also included those aiming to identify differences in flipped and traditional pedagogic approaches, such as “What do you think of a flipped/traditional classroom?” and “What differences do you think there are between flipped and traditional classrooms?”

Research procedure

The study adopted a pretest—treatment—posttest design. Before the treatment, the questionnaire was ensured internally consistent in order to improve the reliability of the result through a pilot study. A business English writing test was also conducted through a pilot study to confirm the business English writing skills of participants were on the same level.

During the treatment phase, Class A received the flipped teaching model, while Class B was taught through the traditional pedagogy. Both classes were taught by an experienced teacher. The teaching period endured for one semester (16 weeks). At the initial stage of the semester, both classes received the pretest in order to determine their writing skills, followed by the treatment phase where one class received the flipped teaching model, while the other was under traditional pedagogy for one semester. Both classes were tested using the above-mentioned writing task and the satisfaction scale to identify the academic achievements and satisfaction levels. The tests were followed by interviews with the participants directed by the researchers.

The clicker aided flipped classroom

The flipped treatment of the research required students to learn and internalize the writing skills on their own prior to physical attendance to the class. They should also complete the business English writing practice before class. In case they have any difficulty or problem, they could communicate with peers or the teacher through online communication tools. Lecture videos were also provided, containing the contents delivered in the traditional classroom. Lecture notes, review comments, and abundant related materials were available in order to facilitate students’ learning before class.

Clicker was used to make the classroom dynamic in the flipped classroom. Students were required to answer questions raised by the teacher in class through anonymous polling, which was realized through the clicker system. Students were also required to actively participate in learning activities, group discussion, idea sharing, problem solution and presentation. The teacher supervised the class and organized activities. When necessary, the teacher would explain difficult questions, present the writing skills and join the discussion.
Clicker-aided polling and group discussion was effective to improve students’ writing skills through scientific arrangements. Firstly, the teacher raised a question regarding business English writing. An example was that students were required to poll for the following sentence completion:

we shall appreciate it if you will effect shipment as soon as possible, thus _____ our buyers to catch the brisk demand at the beginning of the season. A. enable; B. enabling; C. enables.

The correct choice is B. enabling. If over 85% students made the right choice, the teacher would move on to the next topic. On the other hand, if over 85% students made a wrong choice, the teacher would pause to further explain the knowledge until 85% students could understand. If approximately 50% students made a wrong or right choice, the teacher would divide them into several groups for further discussion, after which they would be required to poll again. The teacher would not move on unless over 85% students perceived the question. The same procedure was applied to the sentence matching, sentence organization, translation, and writing.

After class, students could go further into writing development by completing writing task, uploading their assignment, and reviewing assignment for each other. The teacher would supervise and encourage students’ online participation by using communication technologies. The frequency of online participation, duration of online learning, and specific online activities would be recorded for the reference of final evaluation.

In the flipped classroom, students owned the possibilities to show themselves, to develop their own learning potentials and to follow their own learning preferences. The flipped classroom is thus student-centered.

The traditional classroom

Under the traditional pedagogic model, students should preview what would be taught in class, followed by physical attendance to class, where the teacher delivered the contents aided by traditional multimedia projecting technology. Students listened to the teacher and answered questions when required. The teacher assigned homework to students and reviewed the homework. In-class activities were limited because the teacher delivery mainly occupied the class. Students had no choice but to follow the teacher’s pace. Therefore, the traditional pedagogy is a teacher-centered model.

Results

The questionnaire in the pilot study

Participants were required to fill in the questionnaire and the data were entered into the computer to analyze the internal consistency, which was shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Scale mean if item deleted</th>
<th>Scale variance if item deleted</th>
<th>Corrected item-total correlation</th>
<th>Squared multiple correlation</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>54.31</td>
<td>119.928</td>
<td>.816</td>
<td>.666</td>
<td>.894</td>
</tr>
<tr>
<td>Class B</td>
<td>55.20</td>
<td>91.576</td>
<td>.816</td>
<td>.666</td>
<td></td>
</tr>
</tbody>
</table>

Cronbach’s Alpha is considered as a correlation that identifies the same construct. The average correlation of a set of items is an accurate estimate of the average correlation of all items that pertain to a certain construct (Nunnally, 1978). As shown in Table 2, the Cronbach’s Alpha is .894, having reached a satisfactory level. This indicates that the items in the questionnaire refer to the same construct. In other words, the questionnaire used in this study is internally consistent.
The pretest in the pilot study

In order to identify the differences in initial business English writing skills between both classes, a pretest was administered to participants, whose result was analyzed in Table 3.

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
<th>Std. error mean</th>
<th>Mean difference</th>
<th>t</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class A</td>
<td>35</td>
<td>70.29</td>
<td>4.170</td>
<td>.705</td>
<td>.119</td>
<td>.112</td>
<td>.911</td>
</tr>
<tr>
<td>Class B</td>
<td>36</td>
<td>70.17</td>
<td>4.724</td>
<td>.787</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 shows the difference in initial business English writing skills between the flipped class and the traditional one before the treatment. It is clearly found that the mean score in Class A (flipped) is 70.29, which is not significantly different from that in Class B (traditional) \((m = 70.17, t = .112, p = .911)\). This indicates that both classes are on the same level of business English writing skills, paving a solid ground for further comparison after the treatment.

The test of research hypotheses in the study

This part will summarize the research results sourcing from the data processed by corresponding programs and it is sequenced based on the research hypotheses.

After one semester’s teaching practice through flipped and traditional approaches in Class A and B respectively, both classes underwent business English writing tests. The results were summarized and analyzed in Table 4.

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent variable</th>
<th>Type III sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected model</td>
<td>Score</td>
<td>90.795(^{a})</td>
<td>1</td>
<td>90.795</td>
<td>4.470</td>
<td>.038</td>
</tr>
<tr>
<td></td>
<td>Satisfaction</td>
<td>872.340(^{b})</td>
<td>1</td>
<td>872.340</td>
<td>9.614</td>
<td>.003</td>
</tr>
<tr>
<td>Intercept</td>
<td>Score</td>
<td>402471.076</td>
<td>1</td>
<td>402471.076</td>
<td>1.981E4</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Satisfaction</td>
<td>198605.974</td>
<td>1</td>
<td>198605.974</td>
<td>2.189E3</td>
<td>.000</td>
</tr>
<tr>
<td>Class</td>
<td>Score</td>
<td>90.795</td>
<td>1</td>
<td>90.795</td>
<td>4.470</td>
<td>.038</td>
</tr>
<tr>
<td></td>
<td>Satisfaction</td>
<td>872.340</td>
<td>1</td>
<td>872.340</td>
<td>9.614</td>
<td>.003</td>
</tr>
<tr>
<td>Error</td>
<td>Score</td>
<td>1401.571</td>
<td>69</td>
<td>20.313</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satisfaction</td>
<td>6260.956</td>
<td>69</td>
<td>90.738</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Score</td>
<td>403873.000</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satisfaction</td>
<td>205408.000</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected total</td>
<td>Score</td>
<td>1492.366</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satisfaction</td>
<td>7133.296</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. \(^{a}\)R Squared = .061 (Adjusted R Squared = .047); \(^{b}\)R Squared = .122 (Adjusted R Squared = .110).

Table 5. Comparison of means between flipped and traditional classrooms

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
<th>Mean difference</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flipped</td>
<td>35</td>
<td>76.43</td>
<td>3.980</td>
<td>2.26</td>
<td>2.11</td>
<td>.038</td>
</tr>
<tr>
<td>Traditional</td>
<td>36</td>
<td>74.17</td>
<td>4.966</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flipped</td>
<td>35</td>
<td>56.40</td>
<td>9.290</td>
<td>7.01</td>
<td>3.10</td>
<td>.003</td>
</tr>
<tr>
<td>Traditional</td>
<td>36</td>
<td>49.39</td>
<td>9.749</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 4, the main effect of the type of class is statistically significant at the significance level .05, indicating that different types of classrooms exerted statistically significant influence on both academic achievements and satisfaction levels. In order to test both hypotheses, comparison of means was necessary, which was analyzed in Table 5.

All the data in the above tables are conducive to arriving at conclusions of both hypotheses.
Hypothesis 1: The flipped business English writing classroom will bring about better academic achievements than the traditional one

Table 5 presents the comparative results between flipped and traditional classrooms. As shown in Table 5, the mean score under flipped classroom is 76.43, which is significantly larger than that under the traditional approach ($m = 74.17$, $t = 2.11$, $p = .038$). This means that under the flipped classroom, students obtained significantly greater achievements than under the traditional one. In addition, as shown in Table 4, class type has significant impact on academic achievements. Therefore, the hypothesis “The flipped business English writing classroom will bring about better academic achievements than the traditional one” is accepted.

Hypothesis 2: The flipped business English writing classroom will be more satisfactory than the traditional one

The comparison of satisfaction levels of different models of teaching is also shown in Table 5. As presented in Table 5, the satisfaction level obtained through the flipped classroom is significantly ($p = .003$) higher than that ($m = 49.39$) through the traditional approach. Furthermore, as demonstrated in Table 4, the type of class has statistically significant influence on the satisfaction level. Therefore, the hypothesis “The flipped business English writing classroom will be more satisfactory than the traditional one” is also accepted.

Results of the interview

The results coming from the interview echoed the acceptance of hypotheses. A high percentage of students (83.27%) deemed the flipped classroom as a more effective approach than the traditional, where they considered that they could obtain better academic achievements in the flipped classroom than in the traditional one. Moreover, the majority of students (81.63%) were satisfied with the flipped model. They tended to enjoy the independent learning because they could arrange their learning time and places freely. The anonymous polling and peer discussion relaxed them and broadened their horizons. The flipped in-class activities were dynamic with the aid of clickers. The abundant online videos and other materials also contented students in that they could learn at their own convenience. The online environment under the flipped model also enhanced their self-efficacy and self-regulation (67.32%).

Discussion

The discussion part is composed of perception of both satisfaction and academic achievements.

Satisfaction

The flipped classroom, a student-centered model, positively influenced students’ psychological state and hence satisfied students. Under the flipped model, students could practice writing at home, in the office, in the classroom or any other place they felt convenient. Students could also learn at any time they preferred. This great convenience undoubtedly satisfied them. However, in the traditional class, students had to attend the physical class at stipulated time strictly according to the preset timetable. They should also complete the homework within stipulated time. Any delay or miss would be punished. Students could only take in what the teacher delivered in class without any self-directed creation. This might have dismayed students and caused dissatisfaction.

Learner-instructor interaction, learner-content interaction, and Internet self-efficacy were good predictors of student satisfaction, and time spent online per week seemed to have influence on peer interaction, Internet self-efficacy, and self-regulation (Kuo et al., 2013). In the flipped classroom, learners were encouraged to interact with the instructor, and they were also required to learn the contents prior to physical class. After class, learners were also urged to interact with peers and the instructor frequently. Recording and evaluating students’ online activities undoubtedly prolonged the time spent on the Internet. The self-directed learning before class might have gradually increased students’ self-efficacy and enhanced their self-regulation as well.
On the contrary, the traditional class failed to promote the peer interaction, and the interaction between peers and the instructor was also discouraged since merely the instructor delivered the knowledge and what students should do was only to follow. The main interaction between learners and contents was also limited to the in-class environment, where students had contact with the contents under the guide of the teacher. After class, students' responsibility was mainly to finish the assignment. The online learning was thus minimized, followed by lower Internet self-efficacy and self-regulation. All of this might have contributed to lower satisfaction levels under the traditional model.

Close collaboration within a team was fairly satisfactory to its members (Napier & Johnson, 2007), because students in teams were highly satisfied with peer interaction (Oncu & Ozdilek, 2013). In a team, students built up the same identity and voiced their opinions so as to consider peers as friends in the same class and mutual trust was therefore constructed, which enhanced the collaboration and mutual support among peers (Finegold & Cooke, 2006). However, collaboration among peers did not ensure a high satisfaction level. Collecting the right people, organizing a proper learning activity and developing the cooperative sense played an important role in satisfaction and success (Kozlowski & Ilgen, 2007), where students were highly motivated (Koh, Wang, Tan, Liu, & Ee, 2009).

In the flipped classroom, students were gathered to participate in learning activities organized by the teacher. The teacher raised some intractable questions for them to deal with together. Students had to collaborate and cooperate in order to successfully tackle them. In the business English writing class, the teacher required students to compose a paragraph in groups. The paragraph was harsh for a single student but easy for a group because a number of proper nouns and technical words were needed. In this way, collaboration was formed, contributing to satisfaction with the class. Students also had the freedom to organize teams on their own, which was helpful for them to collect right learners, improving collaboration and mutual trust. The important matter for the teacher to cope with was to organize the activities that could attract and activate students.

**Academic achievements**

Past literature revealed that learner satisfaction in online learning was closely related to persistence and lower dropout rates (Arbaugh, 2000; Billings, 2000; Levy, 2007; Thurmond, Wambach, Connors, & Frey, 2002), which was an essential element leading to high academic achievements (Workgroup, 1997; Biner, Dean, & Mellinger, 1994; Chang & Smith, 2008; Marks, Sibley, & Arbaugh, 2005). The more satisfactory the online learning was, the higher academic achievements the learners would attain to (Puzziferro, 2008). Therefore, an online course was successful when students were satisfied with their learning experience (Marks et al., 2005) and students experienced success in learning the course content (Chang & Smith, 2008; Marks et al., 2005; Puzziferro, 2008). The flipped pedagogic approach included intense online learning before and after class. Students’ satisfaction with the online learning decreased dropout rates, enhanced students’ learning persistence and attracted more efforts in learning, hence they achieved success. It is therefore reasonable for students to obtain higher academic achievements under the flipped model than the non-flipped in this study.

A strong positive relationship was found between cognitive map improvement and academic achievements. Students with richer, and more complicated cognitive maps possessed more advanced mechanisms, which deepened their perception of learning, leading to higher academic achievements (Jones et al., 2014). In this study, through online learning, peer discussion, self-directed learning and other factors under the flipped model, students made less mental efforts and the amount of cognitive engagement was reduced (Clark et al., 2006) by successfully managing cognitive load (Abeysekera & Dawson, 2015), possibly resulting from more complicated and suitable cognitive maps, which was helpful for better academic achievements.

However, in the traditional classroom, students had no choice but to follow what the teacher delivered. They were not encouraged to form their own cognitive maps. Instead, most of the cognitive maps were unanimous resulting from the sole delivery from the teacher. Lack of plentiful peer interaction and online learning also hindered the formation of advanced cognitive maps, which definitely caused lower academic achievements.

It was indicated that there were significantly positive relationships between self-regulated online learning and academic achievements (Lim & Morris, 2009). In other words, the stronger the self-regulated learning motivation was, the better academic achievements students would obtain (Artino, 2009; Artino & McCoach, 2008; Paechter et al., 2010; Puzziferro, 2008; Yükselturk & Bulut, 2007). Before and after the flipped class, students were required to
share ideas with peers and the teacher online, and the online participation was also saved as final performance assessment. Students should prepare themselves for online activities. Otherwise, they could not successfully join the online interaction. Nevertheless, it was enough for students to preview what would be taught and review what had been learnt in the traditional teaching approach. Consequently, the flipped model required students to engage in more self-regulated learning activities than the traditional. It was thus logical to arrive at the conclusion that the flipped model contributed to better academic achievements than the traditional.

Students with stronger self-efficacy toward online learning tended to be more motivated and thus obtained better academic achievements (Wang et al., 2013). Self-efficacy was defined as “the beliefs in one’s capabilities to organize and execute the courses of action required to manage prospective situation” (Bandura, 1995, p. 2). In other words, self-efficacy is the belief held by learners that they can succeed in a specific subject or task. Self-efficacy exerted a great influence on motivation and academic achievements (Schunk & Pajares, 2002). In the flipped pedagogic approach, students could determine the learning time before and after class on their own until they felt they were enough informed. They were provided with abundant learning materials such as lecture videos, lecture notes and peer opinions. In this way, they tended to be full of confidence before they entered the classroom to share with peers. On the contrary, the traditional model allowed students to complete assignment and preview the contents before class. No enough time was left over for them to conduct self-directed learning. The higher achieving students might be able to keep pace with the teacher, while the lower or intermediate achieving students might fail. As a result, they might possess lower levels of self-efficacy. Consequently, higher self-efficacy in the flipped approach contributed to better academic achievements than the traditional one.

Except for the self-efficacy, the ability to use online technologies, such as emailing, typing, uploading, downloading, distance communication, and Internet exploring, also greatly influenced the effectiveness of online learning (Wang, et al., 2013), hence the flipped model was under the impact of student online technology mastery, which was positively correlated with online academic achievements (Joo, Bong, & Choi, 2000). In other words, the more skillful use of online technology, the better academic achievements would be brought about. Online technology was reported to be familiar to all the participants in this study. This was the indicator causing higher academic achievements in the flipped classroom than in the traditional one because online learning was much more frequently adopted in the flipped than in the traditional.

Use of clickers in the flipped classroom enabled students to obtain positive academic achievements. In the flipped classroom, students polled to choose the answer anonymously by clicking buttons. All the choices were projected onto a large screen through the multimedia projecting equipment. Choices were presented through statistically analyzed histograms so that the frequency and proportion of the choices could be easily found. The teacher could also identify the number of students who made the right or wrong choices. In case the majority of students were right, the teacher would move on to the next topic. On the contrary, if the majority were wrong, the teacher would explain the contents again before moving on. The teacher also left time for students to discuss with peers before making the decision. Students could also communicate with the teacher if stuck in trouble. In this way, clickers activated the teacher to break the monotony, increased the interaction between the teacher and students, and realized the immediate anonymous feedbacks (Arnesen et al., 2013). The learning atmosphere was therefore stimulated and the dynamic interaction and peer discussion could produce many new ideas in a relaxed manner since students could anonymously voice their opinions. Academic achievements were also realized due to more participation and growing interest in learning.

Conclusion

Advantages

This study adopted both qualitative and quantitative research methods to enhance the reliability of the results. Participants were randomly selected and the number was also proper to represent the population. The research instruments, such as the business English writing test, the questionnaire and the interview, were valid and internally consistent. The research procedure was also scientifically administered involving the pre-and-post tests and the treatment. The statistical comparison was also well based on the prior foundation, such as the same business English writing skills and the qualified teaching.
Although a number of studies explored the effectiveness of flipped class (e.g., Asef-Vaziri, 2015; Butgereit & Osman, 2014; Flumerfelt & Green, 2013), this study is obviously different from them. The study integrated the popular educational technology, i.e., clickers system, into the flipped class, which has been hardly examined by previous studies. However, if we had applied our technologies to other more specific problems of the flipped classroom not studied before, we could have produced more interesting and innovative results. Such problems could be to study, for example, the influence of specific learning material and/or learning activities specifically designed for the flipped classroom on students’ engagement, motivation, peer interaction and self-regulation.

**Disadvantages**

Nevertheless, disadvantages could not be avoided. Writing skills are quite elusive and hard to determine. Although this study scored the business writing tests based on the criteria, misjudge was difficult to be eradicated. Despite that both classes were taught by the same teacher, the teacher might have performed differently in classes of different styles.

The traditional class in this study was chosen to follow a very strict behaviorist model which, by default, will lead to worse results compared to the flipped model. As such, the prediction of the two research hypotheses/questions is rather obvious, since the traditional class B has really very poor chances to perform well. In the real life, a traditional class model is not so strongly behaviorist and may include some constructivist elements.

The problem addressed by the paper concerns the better academic achievements and satisfaction experienced by students in a flipped classroom about business English writing. This same problem has been extensively explored by a lot of other studies carried out in other disciplines or topics. This fact reduces the significance of the problem since the results are expected to be very much the same as in the current research work. The paper could have gained more significant value if it had offered alternative points of view by focusing on more specific problems and issues rather than addressing the same generic problem as previous studies.

Furthermore, normally in a traditional classroom, students’ writing task is often completed at home, and then the teacher will give her or his evaluation on students’ solutions in the next class. This culture covers a part of the flipped model that was mentioned in the paper.

**Directions in future research**

Future studies could pivot on different courses and expand the research scopes related to writing skills development. Business English writing skills might not be able to fully determine whether the flipped classroom is effective. More courses such as mathematics, chemistry, physics and those in higher education might also be examined. More detailed aspects of writing could also be explored. Examples are paragraph composing, lexical collocation and sentence organization which could be strong indicators of writing skills. Merely global scoring of writing might have misled the results. In addition, for business English writing, the clickers might not have larger space or contribution to be applied, thus the teachers maybe have to spend time designing the activities to facilitate this kind of writing with peer or curricula supports.

The clicker used in this study plays a significant role of facilitating the flipped classroom. The flipped classroom usually requires student’s readiness for discussion in the classroom, rather than one-way lecturing. The clicker-assisted anonymous polling and peer discussion greatly encourage students to participate in the flipped class. So the clickers may be one good facilitating mechanism in the classroom. However, the teacher requires some strategies or scaffolding to promote discussion in class and readiness for discussion before class. Future studies on the flipped classroom may also highlight the strategies, scaffolding, or educational technologies to promote both in-class and out-of-class discussions.
Acknowledgments

The author wishes to thank the anonymous reviewers, the people who help this study and the projects which financially support this study: Findings of 2014 Research Project of Philosophy and Social Sciences of Zhejiang Province “Use of Clickers in College English Class in Zhejiang Province” (14NDJC245YB), 2013 Philosophy and Social Science Guidance Research Project of Education Bureau of Jiangsu Province (Project No.: 2013SJJD740005), Special Fundamental Research Fund for the Central Universities (Project No.: 2013B33914), Special Fundamental Research Fund for the Central Universities (Project No.: 2014B10514), 2013 Shaoxing Important Research Project of Higher Education Reform, and 2014 Research Project of Zhejiang Yuexiu University of Foreign Languages (Project No.: N2014013), 2014 Shaoxing Higher Education Excellent Course (undergraduate) “Business English Writing,” 2015 Excellent academic monograph publication fund of Zhejiang Yuexiu University of Foreign Languages.

References


Triantafyllou, E., & Timcenko, O. (2014). *Introducing a flipped classroom for a statistics course: A Case study*. In *Proceedings of 25th Annual Conference European Association for Education in Electrical and Information Engineering (EAEEIE)* (pp. 5-8). doi:10.1109/EAAEIE.2014.6879373


### Appendix A. A satisfaction scale (adapted from Stokes, 2001)

<table>
<thead>
<tr>
<th>N</th>
<th>Questions</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I am able to access a computer with an Internet connection to do my work for the flipped class.</td>
<td>5=very often; 4=often; 3=sometimes; 2=seldom; 1=never</td>
</tr>
<tr>
<td>2</td>
<td>The resources I need for the flipped class are readily available through the Internet.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>I am satisfied with the degree of contact I have with my teacher when working through the flipped class.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I am pleased with the success I am having with completing the flipped class.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>My technology knowledge level is sufficient for learning in a flipped environment.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I am feeling somewhat connected to the University setting by taking a class that places emphasis on learning through the flipped classroom.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I would prefer to take more of my classes through Internet delivery.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Participating in a flipped class has allowed me more flexibility in my daily activities.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I would prefer more of the course materials in my traditional face-to-face classes to be in a flipped format.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I believe that working in a flipped environment enables me to play a more active role in the learning process.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Communication with other students through Internet is a positive experience.</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>I find the online tutorials to be useful in helping me understand the material.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>My flipped class is providing me with skills that I can use in other courses.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>I believe that the flipped environment is preparing me for future profession development.</td>
<td></td>
</tr>
</tbody>
</table>
A Review of Research on Intercultural Learning through Computer-Based Digital Technologies

Emrullah Yasin Çiftçi
Middle East Technical University, Faculty of Education, Department of Foreign Language Education, Ankara, Turkey // yciftci@metu.edu.tr

(Submitted March 22, 2015 Revised June 20, 2015 Accepted August 10, 2015)

ABSTRACT

Intercultural communication is now a crucial part of our globalizing lives; however, not everyone has an opportunity to engage in an intercultural interaction with people from different cultures. Computer-based technologies are promising in creating environments for people to communicate with people from diverse cultures. This qualitative synthesis of quantitative and qualitative research therefore aimed to investigate the literature in respect to intercultural learning through technology use. Besides reporting research descriptives in the literature, this review brought eight main points to the light: (1) an overall satisfaction with digital tools and intercultural learning, (2) increased knowledge toward both own and target culture, (3) superficial and fact-based exchange between similar cultures and profiles, (4) varied levels of intercultural communicative competence development, (5) lack of in-depth analysis and of detailed reports, (6) necessity for training, guidance, and good communication skills, (7) need for stimulating contexts, and (8) technical and institutional challenges. The current study is timely to help researchers and practitioners to both comprehend the existing literature and to find new research and practice directions in the field.

Keywords

Intercultural learning, Intercultural communicative competence, Intercultural exchange, Telecollaboration, Computer-mediated communication

Introduction

Intercultural interactions and exchanges have started to gain an important place in our everyday life as one of the indirect results of our globalizing world. This state of interconnectedness among people and cultures has been so apparent that it is claimed nowadays that “international relations become the basis for a well-functioning tomorrow.” (Schenker, 2012, p. 449). In line with the impact of global developments on functioning of modern societies, having global communication networks and developing an intercultural communicative competence (ICC) have become one of the aims of today’s educational systems (Alptekin, 2002; Thorne, 2003). Thus, ICC is strongly supported as a part of language instruction in schools as it enables people to communicate and negotiate successfully with people from other cultures (Byram, 1997; 2000).

Although no agreement has been established on the definition of ICC, there seems a general agreement on its key goals (Deardorff, 2006; O’Dowd, 2003). Byram’s (1997) ICC model, which is the most commonly referred one in the literature, covers attitudes, knowledge, skills of interpreting and relating, skills of discovery and interaction, and critical awareness. Within his ICC model, the aim is to go beyond superficial intercultural fact exchanges and to develop a deep intercultural learning. The model suggests people to be open-minded, reflective on cultural similarities and differences, and to reduce their ethnocentrism. Developing this intercultural competence seems as a challenging task but is necessary to be a part of the global world (Talkington, Lengel, & Byram, 2004). Teachers of foreign languages or of other relevant fields therefore need to find sound ways to integrate intercultural competence into their courses. Moreover, cultural knowledge alone may not precede intercultural competence (Bennett, 2009; Perry & Southwell, 2011; Talkington, Lengel, & Byram, 2004), so a critical examination of cultures is a must to go beyond simple facts and knowledge about cultures, which is possible through interactions with people from other cultures.

With respect to helping every student to develop an intercultural competence, computer-mediated communication (CMC) and telecollaborative projects provide environments for students to have an opportunity to communicate with people from diverse cultures (Belz & Thorne, 2006; Kern, 2006; Lee, 2009; Müller-Hartmann, 2000; O’Dowd, 2007). Telecollaboration studies in fact have started to gain more importance due to their practical nature in terms of intercultural contacts. However, existing studies stress a lack of research in technology-oriented intercultural learning.
environments (Jin & Erben, 2007; Lee & Markey, 2014; O’Dowd, 2007; Perry & Southwell, 2011). Perry and Southwell (2011) especially state, in their review of intercultural models and approaches, that there is no thorough research on the development of intercultural competence via digital technologies, so, according to them, the area needs further research. To my best knowledge, there is no review research on this issue as well. Thus, this review study aims to fill a gap in the current research in intercultural learning through computer-based technology use in order to mark prevalent issues in the field. These major questions therefore seek for answers in this review study:

- What kinds of technologies were used in intercultural studies?
- What type of participants and contexts were involved?
- How long did the studies take place?
- What were the major findings in terms of intercultural learning?
- How effective were digital technologies to promote intercultural learning?
- Are there any potential gaps and suggestions for further research directions?

**Research methods**

Different inquiry styles utilizing quantitative and qualitative paradigms in this technology-oriented intercultural field made it more appropriate to have a qualitative synthesis of quantitative and qualitative research (Baran, 2014; Suri & Clarke, 2009). As it has been implied, this review did not adopt a positivist approach to synthesis; rather, it benefited from the interpretative and transparent nature of qualitative synthesis (Suri & Clarke, 2009). The study also did not follow an existing model or methodology but it was inspired and directed by Baran (2014), Cooper and Hedges (2009), and Suri and Clarke (2008). The systematic steps that were broadly followed were problem formulation, literature search, data evaluation, data analysis, and interpretation of the results (Cooper & Hedges, 2009).

![Flow chart of literature search results and selection criteria](image)
Articles for this synthesis were chosen following a systematic process of selection (see Figure 1 for a summary of the selection process). As a result of the inquiry, 26 studies were identified for the review (Appendix A). Time frame was set as 2004 and 2014 covering a 10-year-long period including recent developments in web 2.0 tools and telecollaboration studies (Guth & Thomas, 2010). In this way, the focus is more on the effect of recent computer-based digital technologies and their effectiveness on intercultural learning, which may pave the way for a more reasonable and innovative research direction.

After identifying the articles, they were examined and coded in an analytic synthesis table (see Appendix A for a detailed illustration of the categories). In addition to this analytic table, major outcomes of the studies in terms of intercultural learning were included as data. The table helped the study to recognize research descriptives in the field while the major outcomes of the studies formed the qualitative findings of this review study after they were coded in a qualitatively emergent/evolving and interpretative fashion (Creswell, 2013).

Findings

Following an open coding process (Elo & Kyngäs, 2008) of major findings and outcomes of the studies (see Appendix B for a coding sample), and collecting an expert opinion on the process of theme formation, several main categories emerged from the selected 26 empirical studies. They are given as (1) research descriptives, (2) an overall satisfaction with digital tools and intercultural learning, (3) increased knowledge toward both own and target culture, (4) superficial and fact-based exchange between similar cultures and profiles, (5) varied levels of ICC development, (6) lack of in-depth analysis and of detailed reports, (7) necessity for training, guidance, and good communication skills, (8) need for stimulating contexts, and (9) technical and institutional challenges.

Research descriptives

Applied technology types and environments created

Most of the reviewed studies did not use only one digital tool in fostering intercultural interactions. In order to have a glimpse of what kind of technologies were used and how frequently they were applied, a frequency graph is provided in Figure 2.
Online discussion boards were the most frequently used ones in the studies. Text-based chat, blogs, and email exchanges followed online message boards in terms of frequency of use. These tools offer both asynchronous and synchronous communication, which reflects the effort put by the studies to take advantage of different types of communication ways to help their participants to develop an intact intercultural competence. Other tools such as video recording, video conferencing, and podcasting were also used to enrich both the nature and function of the communication. Microblogging -Twitter- was used only once in the study conducted by Lee and Markey (2014).

Study contexts, countries and subjects

Studies were conducted with participants from at least two different countries or cultures. Except three studies, all of the exchanges took place only on online platforms. As for exceptions, Lee (2011) conducted the research in a study abroad context where participants had a chance to have a face-to-face interaction with the target culture, so technology was used to enhance their intercultural development in that context. Elola and Oskoz (2008) created an online environment where American students in a study abroad environment maintained communication with the students at home university to increase home students’ intercultural understanding. Canto, Jauregi, and van den Bergh (2013) included an experimental group that explored intercultural aspects only in inauthentic classroom environment, and they compared this group with an online interaction group in which language learners had a chance to communicate with native speakers.

Regarding the countries and interactions involved in the studies, Spain-USA exchange \((n = 4)\) formed 15.3% of all intercultural exchanges. Following this pair, there were Germany-USA \((n = 3)\), and Taiwan-USA \((n = 3)\) exchanges. USA-China \((n = 2)\) and Chile-USA \((n = 2)\) followed these pairs in terms of frequency. Lastly, there were other interactions that were established only once. You can refer to Appendix A to see these remaining exchanges. However, there were four studies which did not require participants to communicate with another culture directly over some technological tools. Lee (2011; 2012) conducted her studies only in Spain with American study abroad students. Ducate and Lomicka (2008) studied only American students in the USA who followed blogs by native speakers of German and French, and wrote their own in time. Lastly, in Elola and Oskoz (2008), American students in Spain helped the students at home university to increase their intercultural understanding through blogs. Overall Figure 3 below illustrates country context of the participants and how many times they were involved in intercultural communication or intercultural learning. The figure shows that USA is the most popular country for technology-oriented intercultural studies.

As for the participants of the studies, it is possible to say almost all studies were conducted between undergraduate students from different cultures even though there were some exceptions that involved in-service teachers (Angelova & Zhao, 2014). MA students (Hauck, 2007) or local people (Lee, 2011; 2012). Language learning or practice was highly frequent in the studies, so sometimes the focus of the studies was not only intercultural learning.
Study duration

Except four studies (Ducate & Lomicka, 2008; O’Dowd, 2005; O’Dowd, 2007; Ware & Kramsch, 2005), studies were conducted around one semester. Since data were mainly collected from undergraduate students at different universities, one semester seemed as an ideal duration for the studies. However, O’Dowd’s (2005; 2007) studies were exceptions, which lasted for nine months and two years respectively, but these were large scale projects designed for longer periods. Moreover, Ware and Kramsch’s (2005) research took only three weeks to complete, which seemed also as an exception to the research descriptives.

An overall satisfaction with digital tools and intercultural learning

Analyses revealed that most of the participants completed intercultural projects with satisfactory feelings although there were some varying voices among participants. Participants generally ended up with a feeling of enjoyment as a result of using technology to establish some types of communication with native speakers of their target language or with people from another culture (Jin & Erben, 2007; Keranen & Bayyurt, 2006; Lee, 2009; Lee, 2012; Lee & Markey, 2014; Liaw, 2006; Liaw & Bunn-Le Master, 2010). Some even reported their emerging desire to study abroad after they engaged in an intercultural exchange (Tudini, 2007; Zeiss & Isabelli-Garcia, 2005). It can therefore be said that CMC played a useful role in providing enjoyable intercultural experiences and in increasing the motivation and preparation for study abroad programs (Angelova & Zhao, 2014; Hauck, 2007).

Digital tools such as blogging, podcasting, email exchanges, and chat rooms were welcomed by the participants and were reported to be valuable and enjoyable in terms of cultural exchanges (Lee, 2011; Lee, 2012; Lee & Markey, 2014; Schenker, 2012; Tudini, 2007). Blogs particularly helped people to create a sense of community (Lee, 2011), or to have flexibility in what to share with other people (Lee & Markey, 2014). Furthermore, blogs sometimes functioned as a window to the target culture and let people have a taste of the target culture (Ducate & Lomicka, 2008). Overall blogs can be regarded as a convenient digital tool to foster intercultural learning. Although they were not implemented frequently, email exchanges and chat rooms were also reported to be suitable and stimulating for intercultural communication (Chun, 2011; Furcsa, 2009; Tudini, 2007).

Increased knowledge toward both own and target culture

Myriad intercultural opportunities were provided by the studies in terms of learning target culture and language (Lee, 2011; Lee & Markey, 2014), and these opportunities contributed to cultural awareness (O’Dowd, 2007) or to intercultural understanding of the participants (Menard-Warwick, Heredia-Herrera, & Palmer, 2013). Since Bennett (1993) and Kramsch (1993) highlight the importance of reflecting on cultural differences and similarities as a crucial and a triggering aspect of intercultural contacts, it can be said that CMC helped people to reflect on both own and target culture, and contributed to the process toward being an interculturally competent person (Canto et al., 2013; Angelova & Zhao, 2014; Liaw, 2006; Keranen & Bayyurt, 2006). As well as stimulating people to contrast and compare cultures, CMC also helped participants grow interculturally (Liaw & Bunn-Le Master, 2010) by increasing their knowledge (Ducate & Lomicka, 2008; Keranen & Bayyurt, 2006; Lee, 2009; Schenker, 2012), interest (Jin & Erben, 2007; Liaw, 2006) and awareness (Angelova & Zhao, 2014; Elola & Oskoz, 2008; Rooks, 2008; Ware & Kramsch, 2005; Zeiss & Isabelli-Garcia, 2005) toward the other and own culture. Reflection on both own and target culture did not only increase intercultural understanding but also resulted in an attitude change or a development of a positive attitude toward other cultures to a certain extent (Elola & Oskoz, 2008; Furcsa, 2009; Schuetze, 2008).

Varied levels of ICC development

Byram’s (1997) ICC model is by far the mostly adopted framework for the studies reviewed here. Adopting a complete intercultural framework as ICC seemed to be helpful for researchers to conduct deeper analyses rather than only focusing on fact-based intercultural exchanges.
The main claim by the telecollaborative studies was that ICC could be developed through facilitating communication between different cultures over appropriate technology (Schenker, 2012). To support this idea, O'Dowd (2007) highlighted the potential of telecollaboration in supporting the development of students’ ICC compared to the traditional instruction. Those studies using ICC in interpreting data reported evidences for different levels of ICC. Lee (2012) showed differing attitude among participants toward the host culture but also showed the signs of a growth from ethnocentrism to ethnorelativism. These varied levels of development could also be seen in Liaw (2006; 2007) where English learners who communicated with native speakers showed four levels of ICC. In addition to Liaw’s studies, Jin and Erben (2007) showed a variable process of ICC among their participants. Their development fluctuated but eventually they improved their intercultural sensitivity. This varied and individual aspect of ICC in the studies indicates that learners show signs of being on the way to develop a complex ICC (Schuetze, 2008). Using a more secure phrase like “being on the way” is actually more appropriate with studies reviewed here because they mostly fail to provide detailed reports for their participants and data analysis process, so strong claims as “they have developed ICC” must be avoided due to ICC’s complex and always evolving nature (Byram, 1997).

Although a clear developmental process among participants was not observed, it is fair to say that participants developed a sense of awareness toward viewpoints of other people and their differing perspectives (Furcsa, 2009; Ware & Kramsch, 2005). It is therefore possible to say that intercultural communication through technology triggered learners to develop interculturally; however, this improvement changed from person to person. Although ICC might not be easily measurable due to its dynamic and ever-changing nature, it was showed that it was possible to track major changes in perspectives toward target culture and in reflections on own culture (Elola & Oskoz, 2008).

**Superficial and fact-based exchange between similar cultures and profiles**

Computer-based digital tools, as it has been shown so far, enabled people to communicate with people from other cultures without visiting each other. These interactions also fostered ICC of participants though there were some variability among participants. While explaining the sources of this variation, it was found that cultural understanding was sometimes superficial and fact-based (Lee, 2009; Lee, 2011; Liaw, 2006), but some superficial understandings were developed in consequence of reflecting on both own and target culture (Ducate & Lomicka, 2008). Similarly, Keranen and Bayyurt (2006) admitted interchanges of ideas might seem to be depending on a superficial level; yet, people might build understanding on this superficiality. Furthermore, Liaw and Bunn-Le Master (2010) found that the majority of exchanges of information were fact-based. At this point it could be necessary to engage a facilitator in the dialogue to maximize the intercultural environment in which people could argue over different viewpoints.

On the other hand, studies mostly took place between similar cultures and profiles. Most of the interactions happened with undergraduate students from USA, and other interactions also were realized between similar Western cultures with a few exceptions from Eastern cultures. That is to say, such exchanges between similar cultures and profiles may not yield rich intercultural learning situations. This does not mean that participants were the same in every aspects, but they sometimes failed to create an optimized situation for challenging and fruitful intercultural discussions, which might have fostered a robust development of ICC.

**Lack of in-depth analysis and of detailed reports**

Most of the studies in the literature were designed around language learning or teaching. Thus, it is fair to say that not all studies concentrated on intercultural processes in details. This situation caused studies to report superficial results in terms of intercultural development. This may also be one of the reasons lying behind varied representations of ICC. For instance, Chun (2011) and Schenker (2012) reported varying levels of development; however, there was not much clear information about participants’ ICC development although there was an attempt to track their intercultural development.

This lack of in-depth analysis of data or trying to measure many components at the same time produced a lack of detailed reports. There were studies in fact reporting intercultural sensitivity developed by individual participants as in Jin and Erben (2007) or in Hauck (2007); however, the majority of the literature offered superficial exchanges and analyses in terms of intercultural learning. On the other hand, it was visible that there were some participants in the
studies who were not happy with the projects. For example, Lee (2011) and Lee and Markey (2014) reported the positive results for their technological interventions but the percentages they gave also included an uneasiness among some few other students. What made the reader suspicious was the lack of the report for this dissatisfaction. Similarly, in Lee (2012), 80% found blogs useful for intercultural learning but it is unknown what the rest felt about blog’s integration and its contribution to intercultural learning. Additionally 30% lacked access to computer in the same study but it was not shown how exactly they went through the online intercultural exchange. This silent scream was frequently observed across the literature such as in Ducate and Lomicka (2008), and in Chun (2011) where some participants showed dissatisfaction with online assignments. These are all given to say that the literature may need more elaboration on all of the participants and their experiences in the future.

Necessity for training, guidance and good communication skills

Intercultural interaction through technology has multiple facets as it has been implied so far. Within this complexity, designing a basic online intercultural or cross-cultural telecollaborative environment might not be enough by itself to promote intercultural learning; therefore, the current literature has revealed that there was and will be a need for training participants for technology-based competences or for the required tools before any online exchange takes place (Chun, 2011; Hauck, 2007; Lee & Markey, 2014). However, providing participants with a basic training may not be enough because sometimes it is possible for researchers to realize that the training given is not sufficient for them to maximize the experience (Lee, 2009). Training before the intercultural experience also needs to be designed and implemented carefully by taking unique characteristics of the participating subjects.

Leaving participants alone with people from other cultures may not yield effective communication, which may necessitate a conscious guidance by interculturally competent persons (Angelova & Zhao, 2014; O’Dowd, 2007; Rooks, 2008; Schenker, 2012; Schuetze, 2008). In this sense, Lee (2009; 2011; 2012) suggested a more active, key and facilitating role for instructors. Additionally, O’Dowd (2005; 2007) defined certain roles for teachers to adopt during projects such as organizer, intercultural partner, model and coach, source and resource. O’Dowd (2005) also highlighted an adaptive role for teachers so that teachers should play a constant role in organizing and adapting their guidelines and activities depending on the emerging circumstances. Moreover, Menard-Warwick et al. (2013) and Ware and Kramsch (2005) highly valued the importance of tutor facilitation in their studies. However, there were some opposing voices to the high engagement of instructors in the projects. One argument came from Tudini (2007) who discussed that ICC could develop more in authentic environments than in the places designed and controlled online, so she defended a freer environment for people to interact interculturally. This suggests a moderate involvement of instructors or researchers in the process.

In addition to the key role of facilitators, Rooks (2008) offered a conscious pairing among different cultures because a small problem between them might cast away all the efforts invested for the promotion of intercultural learning. On the other hand, a preliminary trust establishment between teachers and students before a telecollaboration study was offered by Menard-Warwick (2009); in this way, facilitators might ensure exchanges go well and misunderstandings are to be prevented. In addition to guidance, there is also a strong need for everyone involved to have good communication skills in order to handle potential misunderstandings or breakdowns (Elola & Oskoz, 2008; Furcsa, 2009; Hauck, 2007; Lee, 2011).

Need for stimulating contexts

As well as providing training on tools and successful communication, a careful planning should also be considered before implementing an online intercultural project or study. Furcsa (2009) highlights the importance of careful planning and appropriate tasks if instructors aim to create opportunities for productive discussions. It therefore seems that the backbone of successful online intercultural interactions is a meticulous design (Hauck, 2007) or stimulating environments which are intended and organized for the unique characteristics of the participating subjects (Menard-Warwick, 2009).

Designing appropriate environments for a certain profile of people to interact online may not be an easy task to achieve; rather, there are certain steps to consider across the studies such as selection of appropriate technologies and materials, grouping or pairing people wisely, choosing discussion topics carefully, developing ways to handle
problem, and perpetuating the communication. The highlighted aspect of topic choice was visible in the studies conducted by O’Dowd (2005), Schuetze (2008), and Zeiss and Isabelli-Garcia (2005). Additionally, Rooks (2008) supported the inclusion of more enjoyable topics to discuss rather than compulsory writings with restricted options. This flexibility of topics was also suggested by Menard-Warwick (2009) to respond to emergent opportunities for discussions.

With respect to the types of digital tools in the studies, Keranen and Bayyurt (2006) underscored the advantages offered by asynchronous communication such as flexibility and in-depth exchanges. Ware and Kramsch (2005) also supported the advantage of asynchronous communication by presenting some difficulties caused by synchronous communication. However, this does not mean that future researchers should give up using synchronous communication tools; rather, they should combine and diversify them to maximize and optimize the intercultural experiences of their unique subjects.

Technical and institutional challenges

As it was reported in research descriptives, most of the studies took around one semester to be completed. The reason for such a time period was generally given as the difficulty of arranging the projects between institutions from different countries for long time periods because each institution or university had different course requirements or even different time zones (Keranen & Bayyurt, 2006; O’Dowd, 2005; Lee & Markey, 2014). This short duration of studies was mentioned in some studies like Angelova and Zhao (2014), Jin and Erbein (2007), Menard-Warwick et al. (2013), Rooks (2008), and Ware and Kramsch (2005) in an explicit way, but it is not difficult to sense the time limitation or constraint on the studies overall.

It is also known that there were some invisible subjects in the studies who were not quite happy with the online projects. One reason for this dissatisfaction was lack of technical insufficiency (Chun, 2011). For example, Rooks (2008) explained negative reaction of his participants with “logistic failures.” Similarly, Liaw (2006) reported how students found it very frustrating when the system did not work properly. These all suggest another facet of online intercultural learning because placing subjects in an online environment is not sufficient to enable them to engage in a meaningful and productive exchange; sometimes they may simply get distracted by technical breakdowns or get demotivated by short duration of the exchanges.

Concluding remarks

This review on online intercultural learning is well positioned since it has revealed sufficient number of issues to be considered. In terms of research descriptives, it can well be said that studies favored certain contexts, participants and digital tools over others, and technology has attracted a modest but limited number of intercultural researchers. Before further studies are conducted, it is important to revise what this review has brought to light. Figure 4 summarizes all the studies conducted in this field and it does so by benefiting from the interpretative nature of qualitative synthesis.

The literature has showed its own advantages, weaknesses or challenges. Advantages were overall satisfaction with digital tools and intercultural experiences over the Internet or technology, increased knowledge of both own and other cultures, and varied levels of ICC development. As for weaknesses or challenges showed by the literature, most studies tended to report superficial findings without an in-depth analysis, and most of them aimed to measure different aspects of the exchanges at the same time rather than only focusing on intercultural issues, which has left us with scant reports on ICC development. Even though subjects had chances to collaborate, share or discuss, they generally had superficial exchanges, maybe due to their similar backgrounds or to the limited nature of online exchanges. Moreover, experiences of discontent participants were not too visible but still they were sensible, which seems still as a mystery to the literature.
Recommendations for future research and practice

This review offers a number of important recommendations for people who have an intention of utilizing technology in terms of intercultural learning. First of all, the overall picture of the literature (Figure 4) has pushed this study to bring up a number of major necessities. One of them is training of participants for technological tools and good communication skills so that they do not have communication breakdowns or can avoid potential misunderstandings.
without having any problem with the use of technology. Even if they know how to use technology and how to communicate, they still may face technical challenges at their own institutions, so this technical and communicational aspect is crucial to keep in mind before any project is conducted. Furthermore, there could be differences in time zone, academic calendar or course requirement between institutions, so people who plan to take action in this field should not neglect these technical and institutional challenges.

Training participants does not seem to be sufficient for a successful intercultural communication. A stimulating environment should also be designed meticulously to help participants to feel engaged and maintain their interaction with other cultures. While designing the environment, the diversification of the digital tools should be taken into account in parallel with the unique aspects of the participating subjects and contexts. It is also strongly recommended to utilize some new tools in line with the developing technologies; in this sense, video technologies seem to offer significant opportunities in terms of enriching intercultural contacts for future practices and research (Bray, 2010). Such designs also require interculturally competent facilitators who should be knowledgeable in both cultures and should be aware of theoretical background of intercultural issues.

Duration for future studies is also a key issue. So far technical and institutional challenges have forced researchers to keep their studies short but in the future this issue should be revisited by the researchers in order to present robust, detailed, and longitudinal findings to the literature. Although it is difficult to assess ICC thoroughly in a short time and in an online environment, the future researchers can adopt different and mixed data collection tools to increase the likelihood of robust outcomes. In this sense, models or inventories such as Developmental Model of Intercultural Sensitivity (DMIS) developed by Bennett (1993) or Intercultural Development Inventory offered by Hammer, Bennett and Wiseman (2003) could be utilized for future research aims. Also innovative and in-depth qualitative analyses can be helpful to offer rich and thick descriptions of the experiences. Menard-Warwick (2009) actually presented an innovative linguistic analysis, so in the future more innovative studies in terms of data collection and analysis are expected. This is not to tell Byram’s (1997) ICC model has failed to measure what it aimed to measure, but the aim is to bring fresh and in-depth insights into technology-based intercultural learning field. Future researchers therefore should be seeking for in-depth interpretations or intercultural representations gained through digital technologies with the help of theoretical models and tools.

Another prevalent shortcoming in the literature was repeating interactions between similar cultures and age or study groups, which sometimes resulted in poor exchanges. For future practices, it will be a lot better to see exchanges between different study contexts. Interactions between different profiles other than undergraduate students may also provide future research with fresh perspectives. Additionally, dominant aspect of language learning in the studies up to now may have caused researchers to limit themselves; however, future studies with an only intercultural focus and with participants who speak English as a lingua franca may yield more fruitful insights into intercultural issues.

References


*Note.* “∗” References marked with an asterisk and “+” symbol show these articles were analyzed for this review study.
### Appendix A: Analysis table

<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects</th>
<th>Study duration</th>
<th>Technology used</th>
<th>Countries involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>O’Dowd (2005)</td>
<td>An advanced group of English learners Students of Spanish</td>
<td>Nine month</td>
<td>Online message boards</td>
<td>U.S.A. Spain</td>
</tr>
<tr>
<td>Ware &amp; Kramsch (2005)</td>
<td>A learner of German A learner of English</td>
<td>Three weeks</td>
<td>Asynchronous online discussion board, Blackboard</td>
<td>Germany U.S.A.</td>
</tr>
<tr>
<td>Zeiss &amp; Isabelli-Garcia (2005)</td>
<td>US university students engaged in CMC with Mexican university students</td>
<td>Fall 2001</td>
<td>CMC through text-based messages, Electronic bulletin board</td>
<td>U.S.A. Mexico</td>
</tr>
<tr>
<td>Liaw (2006)</td>
<td>Freshmen at a Taiwanese university students from U.S.</td>
<td>One semester</td>
<td>A web-based environment, E-referencing tools, Online forums</td>
<td>Taiwan U.S.A.</td>
</tr>
<tr>
<td>Kerenen &amp; Bayyurt (2006)</td>
<td>In-service English teachers from Mexico Turkish undergraduate students</td>
<td>Not specified in details</td>
<td>Online discussion rooms, Blackboard</td>
<td>Mexico Turkey</td>
</tr>
<tr>
<td>Liaw (2007)</td>
<td>Freshmen at a university in Taiwan American students</td>
<td>One semester</td>
<td>A web-based environment, E-referencing tools, Online forums</td>
<td>Taiwan U.S.A.</td>
</tr>
<tr>
<td>O’Dowd (2007)</td>
<td>German university students Two American, one Irish partner classes</td>
<td>Three qualitative studies between 2001 and 2003</td>
<td>Email, Online message board, Online content materials, Video conferencing Chat rooms</td>
<td>Germany Ireland U.S.A.</td>
</tr>
<tr>
<td>Tudini (2007)</td>
<td>Australian learners of Italian Native speakers of Italian</td>
<td>From April to June 2003</td>
<td></td>
<td>Australia Italy</td>
</tr>
<tr>
<td>Ducate &amp; Lomicka (2008)</td>
<td>German and French language students</td>
<td>Two semesters</td>
<td>Blogger</td>
<td>U.S.A.</td>
</tr>
<tr>
<td>Authors</td>
<td>Current Students</td>
<td>Previous Students</td>
<td>Time Frame</td>
<td>Course Materials</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Rooks (2008)</td>
<td>Japanese university students Thai university students</td>
<td></td>
<td>Six weeks</td>
<td>Email exchanges</td>
</tr>
<tr>
<td>Schuetze (2008)</td>
<td>Students from University of British Columbia and University of Kiel</td>
<td></td>
<td>Six weeks in 2004 and Six weeks in 2005</td>
<td>Email, WebCT</td>
</tr>
<tr>
<td>Furcsa (2009)</td>
<td>Pairs of American and Hungarian undergraduate students</td>
<td></td>
<td>Spring semester of 2005</td>
<td>Email messages</td>
</tr>
<tr>
<td>Lee (2009)</td>
<td>American undergraduate students Spanish undergraduate students</td>
<td></td>
<td>Spring of 2008</td>
<td>Moodle, Blogger, Audacity, iMovie</td>
</tr>
<tr>
<td>Menard-Warwick (2009)</td>
<td>Undergraduate students at a Chilean university Graduate students at a US university</td>
<td></td>
<td>Eight weeks</td>
<td>MSN Messenger</td>
</tr>
<tr>
<td>Liaw &amp; Bunn-Le Master (2010)</td>
<td>Freshman students majoring in English in Taiwan Pre-service teacher education trainees in the U.S.</td>
<td></td>
<td>One semester</td>
<td>A project website called CANDLE</td>
</tr>
<tr>
<td>Chun (2011)</td>
<td>Students learning German at an American university Students studying English at a German university</td>
<td></td>
<td>10-week period</td>
<td>Asynchronous forum discussions, Synchronous text chats</td>
</tr>
<tr>
<td>Lee (2011)</td>
<td>American undergraduate students enrolled at a study abroad program Spanish local people</td>
<td></td>
<td>One semester</td>
<td>Blogger, Blackboard, (Face to face interaction)*</td>
</tr>
<tr>
<td>Lee (2012)</td>
<td>American study abroad students Native Spanish speakers</td>
<td></td>
<td>One semester</td>
<td>Blackboard, Blogger (Face to face interaction)*</td>
</tr>
<tr>
<td>Schenker (2012)</td>
<td>American undergraduate students German high school students</td>
<td></td>
<td>Six weeks</td>
<td>Email exchanges</td>
</tr>
<tr>
<td>Canto, Jauregi, &amp; van den Bergh (2013)</td>
<td>University students forming three groups (1. Virtual worlds with native student</td>
<td></td>
<td>February-April 2010</td>
<td>Adobe-Connect, Second Life, Open Sim (Classroom)</td>
</tr>
</tbody>
</table>
teachers 2. Video-
web communication
with native student
teachers 3.
Nonnative-nonnative
interaction in the
classroom)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Students</th>
<th>Duration</th>
<th>Tools</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menard-Warwick, Heredia-Herrera, &amp; Palmer (2013)</td>
<td>Prospective teachers studying English in Chile</td>
<td>Eight weeks</td>
<td>Online chat groups, MSN Messenger, Moodle</td>
<td>Chile, U.S.A.</td>
</tr>
<tr>
<td>Angelova &amp; Zhao (2014)</td>
<td>U.S. in-service and pre-service teachers Chinese students majoring in English</td>
<td>One semester</td>
<td>Email functions of Blackboard, Chat rooms, Skype (only one pair), Discussion board</td>
<td>U.S.A, China</td>
</tr>
<tr>
<td>Lee &amp; Markey (2014)</td>
<td>Advanced Spanish speakers Advanced English speakers</td>
<td>One semester</td>
<td>Twitter Blogger Audioboo iMovie Movie Maker</td>
<td>Spain, U.S.A.</td>
</tr>
</tbody>
</table>

Appendix B: A Coding sample for major findings

| Lee (2009) | (1) Students reacted positively to the inclusion of the online exchange project and (2) were satisfied with the outcomes. (3) the experience helped them to view cross-cultural learning from a different angle. (3) They reported that they would not have a deeper understanding of certain aspects of Spanish culture if they did not have an opportunity to have a contact with Spanish people. (4) Some found communication short and unyielding but not elaborated on the reasons. (5) Instructors played a key role in facilitating online exchanges. (6) Blogs and (7) podcasting were (8) useful in letting students to engage in exchanging ideas. (9) Blogs also fostered critical thinking and deeper understanding of the topic. (10) More than 70% found online tasks time consuming. (11) In studies like this, they build the environment, group the learners, and observe well but report less, so we do not see how deep intercultural learning was. | (1) Positive reaction (2) Satisfaction with the project (3) Development of cross-cultural understanding (4) Superficial intercultural exchange (5) Necessity for guidance (6) Blogs as a useful tool (7) Podcasting as a useful tool (8) Benefits of combining tools (9) Blogs for deeper understanding (10) Time concern in online tasks (11) Superficial reports (these codes were then color coded and clustered into major themes through a constant comparison among studies. This coding process was done only by the author under the supervision of an expert.) |
Podcasts: A Factor to Improve Iranian EFL Learner’s Self-Regulation Ability and Use of Technology

Somayeh Naseri* and Khalil Motallebzadeh

English Department, Torbat-e-Heydarieh Branch, Islamic Azad University, Torbat-e-Heydarieh, Iran // somaye.naseri@gmail.com // k.motalleb@iautorbat.ac.ir, kmotallebz@gmail.com

*Corresponding author

(Submitted April 15, 2015; Revised June 17, 2015; Accepted July 27, 2015)

ABSTRACT

This study investigated the effect of listening to podcasts on Iranian upper-intermediate EFL learners’ self-regulation ability and their perception toward the use of technology. To meet the objectives of the current study, 54 female Iranian EFL learners were selected. In experimental group they listened to podcast files while in the control group they listened to a radio program. The practices were done through cognitive and metacognitive strategies. A self-regulation ability questionnaire was administered among the participants of both groups. Also, the data obtained from learners’ self-reflection journals were added and interpreted qualitatively and supported the results of this study. Keeping a self-reflection journal provided a detailed account of learners’ thoughts and perception regarding their listening to podcasts. It revealed learners’ positive prospective toward podcasting application in the language learning process and it was clear that the participants in experimental group played a much more dynamic role as a learner. According to the findings of this study use of podcasts as an educational technology can significantly influence and foster self-regulation ability of the learners. Most participants showed a positive perception toward the use of podcast as an educational technology in language learning.

Keywords

Podcast, Self-regulatory ability, Cognitive strategies, Metacognitive strategies, Educational technology

Introduction

Nowadays English language as an international language seems dominated every community. It is not only widespread in formal and academic organizations, but also in everyone’s life around the world. People with different origins are trying to communicate and meet their needs through English language for a variety of purposes. One of the major needs of Iranian EFL learners to employ English in their life is improving their learning skills to learn in what way to learn, be innovative in their learning process and self-regulates the resources, duties, and tasks to set the goals of learning, and accelerate learning. Therefore, one of the foremost abilities to learn is to know how to develop, implement and create a flexible plan in order to obtain the desired outcome in learning the English language. For decades the academic associations were trying to enhance all skills, but still there are some intricacies to create an open learning context. Hence technology is one of the requirements of learners to be employed in today’s life. There are various types of educational technology applications, gadgets, and software that could be useful to be used by the learners. Some of them seem more beneficial to support the learning process. Many researchers have investigated the application of technology by EFL learners not only synchronously in their classroom, but also asynchronously outside the class anywhere at any time.

McGarr (2009) noted that podcasting refers to the sharing audio or video files in digital layout. It can be downloaded manually from the internet or automatically distributed to subscribers. These files are accessible straight from the desktop or transmittable to a media device. The term podcasting came out from the application of Apple’s portable audio player, the iPod. Moreover, other companies created a range of transportable audio and video players in which the term podcasting issued to describe the broadcasting of all audio or video files across the Internet.

Podcasting is a new means employed by teachers accompanying educational content to encourage learning (Hasan & Hoon, 2013). Based on a research done by Ashton-Hay and Brookes (2011) podcasting would enhance learners’ language skills, especially at the field of speaking and listening. Much research findings suggest that podcasts can significantly help learners’ language learning (Hasan & Hoon, 2013). The feasibility of podcasting to teaching English as a kind of educational technology needs researching. This paper explored if listening to podcast has any significant effects on improving self-regulation ability (SRA) of Iranian EFL learners to enhance language learning in the technology era.
(De) limitations of the study

The researcher samples randomly from EFL Iranian learners with upper-intermediate level of language proficiency in Torbat-e-Heydarieh, Khorasan Razavi, Iran. All participants are female adult learners with the same levels of language proficiency and social background. They enrolled at the private language institutions to be prepared for the IELTS test, and the number of the participants has been limited. As the participants are selected from the available sample, there was no control on their age range. Therefore, the participants’ personal learning styles are not controlled in this study. Moreover, the nature or the type of material selected from the podcast listening files might be more interesting than the radio program for the learners, so we cannot generalize the results of this study. Therefore, those who are intended to draw on the findings of this study should be careful with generalizing the outcome.

Literature review

Podcast in language learning

Based on the investigations on podcasting pedagogy, podcast seems a tool to facilitate learners developing English language skills. Hasan and Hoon (2013) has shown that podcasts could support learning not only in speaking and listening, but also in other language skills and areas such as pronunciation, lexicon, and grammar. It also provides opportunities for EFL teachers to simplify learning a language and find a solution for the traditional teaching and learning methods difficulties.

According to a study done by Kargozari and Zarinkamar (2014) in Iran, application of podcasts as an educational device in learning vocabulary was considerably more effective than techniques used traditionally in EFL classes.

Istanto (2011) and Chan, Lee, and McLoughlin (2006) suggested that podcasts can be used to accelerate EFL learners’ proficiency in English as supplementary material to support the course objectives in line with the course materials.

Vandergrift (2006) cited that listening comprehension ability may depend on metacognitive knowledge, for instance record types, listening strategies and objectives. Regarding the researcher experience there is a wide range of podcasts on the web that provides opportunities for every learner to be employed as a supplementary and asynchronous material. It can enhance learners’ listening comprehension ability.

Kan (2010) investigated that Web content designers ought to elucidate their rationale, systematize the contents and regularly enrich its exercises to meet the learners’ need in developing their listening skills.

The impact of listening to podcasts on the listening comprehension ability of some Iranian EFL university students has been investigated. The results of this study proved that participants in the experimental group who utilized podcasts for their listening activities had higher and significant performance than the ones in the control group. They also had positive attitudes toward podcast application for listening. (Ashraf, Noroozi, & Salami, 2011)

Podcasting, authentic material, and technology

Kavaliauskienë (2008) recommended blended learning, which is a combination of multiple approaches to learning via matching online listening with classroom audition activities in teaching or learning English. The results revealed that more practice was required to enhance listening competence. Regarding the findings of this research there are several implications such as: promoting and encouraging learners to improve skills of listening without being frightened by possible failure, growing students' awareness of appropriate individual techniques to accelerate the listening skill to support language learning, the newness and assortment of out-of-class listening to prompt learners to improve their skills without no observation of their peers or teachers, and less nervousness corresponding online listening with classroom audition activities could benefit all learners. Learners will know that listening skills can be improved through a lot of practice. Blended learning also offers a kind of individual self-assessment by a self-evaluation report in personal weblogs, and promotes learners toward higher improvement.
O’Brien and Hegelheimer (2007) found that the use of podcasts is an integration of CALL activities into academic English course regarding the strategies used in listening comprehension. Their study showed positive perspective of both students and teachers toward podcast utilization as an effective element to organize the curriculum of listening courses.

Stanley (as cited in Sze, 2006) described podcasting as a novel technology, which has a high potential to enhance second language learners’ listening and speaking comprehension ability. Since downloading podcasts on different gadgets is easy the learners would be able to involve in many listening practices anywhere. The ease of creating podcasts and high range of audiences around the world motivates students to produce their own podcasts. Moreover, podcast production provides plenty of meaningful language use to the learners.

In an empirical study, Shoar, Abidin, and Pour-Mohammadi (2011) investigated that the students have got higher motivation toward language learning by using mobile. Based on the results of their study technology can provide ample opportunities for learning and generate effective learning. The mobile application can promote teaching English formally at school. They revealed that timing can be accelerated via learning opportunities provided, in an informal setting in their free time. To review the course material by podcasts and mobile devices the performance of the students has been increased after sessions of teaching and before exams, hence the students can also access to the learning materials prior to exams and it can be considered as the result of their better and more frequent access to the learning material during the preparation days before the exam and be encouraged to apply technology as an academic means.

Podcasting has been introduced as a friendly technology in the educational context applied easily by teachers and students at any places or time since most of the educator’s poses one form of gadgets these days including an iPod, laptop, mobiles. Moreover the new teenagers and adults have involved in technological era and there is more material available to them on the web hence the learners are much more eager to utilize this content while learning a language (Mohammadzadeh, 2010).

Self-regulation ability

Investigations on self-regulation initiated in the 1960s and 1970s. The researchers mainly focused on cognitive and metacognitive aspects of human traits to enhance learning by the learners (Schunk & Zimmerman, 2012).

Bandura and his colleagues (as cited in Schunk and Zimmerman, 2012) explored other aspects of self-regulatory processes like social and motivational features. According to Bandura (1991) self-regulation consists of three processes: self-observations (pursuing an activity and selecting the proper strategy to reach the operational stage), self-judgments (comparing one’s action to the norms), and self-reactions (the behavioral and motivational features of a learner).

Self-regulated learning framework

Schraw, Crippen, and Hartley (2006) drew a framework for self-regulated learning; they categorized SRL into three branches including:

- Cognition referring to simple strategies (rehearsal, elaboration, help seeking), problem solving (algorithms, troubleshooting), and critical thinking (identify components, reasoning)
- Metacognition consists of knowledge of cognition (declarative, procedural, conditional) and regulation of cognition (planning, monitoring, evaluation)
- Motivation involves self-efficacy (outcome expectancy, belief, vicarious learning, and modeling) and epistemology (number of beliefs, world view)

Self-regulation strategies

According to Zimmerman and Schunk (2011) social and contextual theories of SRL consider learners in contexts, pay attention to the learners’ opinion, and communication between them. They implied that there are some strategies
used by individual successful learners referring to three main self-regulation strategy classes (behavioral, personal, and environmental).

**Development of academic self-regulation processes**


Arsal (2010) found that the teachers play an important role to develop learners’ self-regulated learning behaviours.

Some other researchers have focused on those learners who self-regulate their activities and performance at school and investigated that those learners obtain a better outcome than the students who are not able to regulate their own practices as cited in Zimmerman and Cleary (2009).

**Research questions**

The research questions posed by the researchers to account for the goal of the study were as follows:

Q1: Does listening to podcasts have any significant effects on Iranian upper-intermediate learners’ self-regulation ability?

Q2. What is the Iranian upper-intermediate learners’ attitudes towards listening to podcasts as an educational technology?

To achieve the appropriate results based on the above questions the following null hypothesis was proposed:

H01: Listening to podcasts does not have any significant effects on Iranian upper-intermediate learners’ self-regulation ability.

**Methodology**

**Participants and setting**

Fifty four female upper-intermediate EFL learners from the Bayan Language Institute in Torbat-e Heydarieh were selected among 118 students. The participants’ age ranged from 14 to 31, but with the same level of language proficiency. Regarding their homogeneity, the institute administered a placement test at the beginning of the semester through which these participants were levelled as upper-intermediate learners. They were divided into two classes as control and experimental groups.

**Instrumentation**

Syndicate (2001) The Second Version of Quick Placement Test developed by Oxford University Press and University of Cambridge Local Examinations Syndicate had been used to choose two nearly homogenous experimental and control groups in which all the participants were at the upper-intermediate level of language proficiency. This placement test has two main parts including part one (Questions 1 – 40) and part two (Questions 41 – 60) that needed to be complete in 30 minutes.

At the beginning of this study a self-regulation questionnaire which was taken from Brown, Miller, and Lawendowski's (1999) was administered between both groups. It includes 63 items on a 5-point Likert scale. The questionnaire has seven subsections: (1) Receiving relevant information, composed of items 1, 8, 15, 22, 29, 36, 43, 50, 57; (2) Evaluating the information and comparing it to norms composed of items 2, 9, 16, 23, 30, 37, 44, 51, 58; (3) Triggering change composed of items 3, 10, 17, 24, 31, 38, 45, 52, 59; (4) Searching for options composed of
items 4, 11, 18, 25, 32, 39, 46, 53, 60; (5) Formulating a plan composed of items 5, 12, 19, 26, 33, 40, 47, 54, 61; (6) Implementing the plan subsections composed of items 6, 13, 20, 27, 34, 41, 48, 55, 62; (7) Assessing the plan’s effectiveness (which recycles to steps 1 and 2) composed of items 7, 14, 21, 28, 35, 42, 49, 56, 63. Test-retest reliability for the total SRQ is high ($r = .94, p < .0001$). Internal consistency of the scale is significant ($\alpha = .91$). After four-month study, the same SRQ was administered in both the control and the experimental groups.

In order to collect some qualitative data on the learners’ thoughts and perceptions toward podcasting as an educational technology, the learners kept a self-reflection journal during the study. Two reflection questions were asked to be answered by each individual as follows:

- **What do you like or dislike about the podcasts?** (You can write about your idea of the content within podcast files, i.e., was the podcast file interesting? If so, why? Was it informative? What have you learned from it? Podcast as an educational technology tool, i.e., what opportunities did it provide that in traditional methods of learning English you did not experience? Or what are some of the activities you did through the use of podcast that you could not do them without it?)
- **Would you express any other ideas or comments related to your experience of listening to podcasts?** Each learner regularly wrote detailed accounts of her attitudes about listening to podcasts in learning the English language as a new technological means and expressed her own voice toward this application and its effect on her performance. Every individual learner recorded her progress and self-evaluation during the course. The key words were entered into the latest version of the SPSS software and were interpreted qualitatively by the researcher. Most participants had positive attitudes towards listening to podcasts and engaged in class activities actively and have gotten less anxiety to participate in the discussion panels related to the files they have listened. They clearly expressed their attitudes toward podcasts as a useful educational technology.

**Procedure**

Prior to the study, the second version of a quick placement test, developed by Oxford University Press and University of Cambridge Local Examinations Syndicate (Syndicate, 2001), was administered to put the subjects in their own levels. Fifty four upper-intermediate female learners were selected among 118 students as the participants of the study. They were assigned into experimental ($N = 26$) and control ($N = 28$) groups. Two weeks prior to the study, the SRAQ administered as the pretest. As podcasting was the treatment of the study, the researcher had chosen 13 files among 23 podcast files as supplementary material to be used asynchronously prior to each session of the class. The podcast files were about personal development and they had been developed by Pavlita (2005). There was no need to listen to the podcast files online but it could be listened asynchronously. The students in the experimental group had the opportunity to choose the time they want to listen. Moreover, the students selected the order of the podcast files that the researcher assigned them to be heard during the term. In the control group the learners were asked to listen to a radio program in English for thirteen times before each session. All activities in both groups were completely the same the only difference was the type of treatment in the experimental group. Based on the teacher’s instruction there was no linear direction to do the activities and students were free to return to any part they needed and follow the process as they have required. The teacher did not teach the content of the podcasts and the students listened to the files prior to their class time. They were used as supplementary materials apart from their main course book. There was just a quick review and evaluation during the course. The key words were entered into the latest version of the SPSS software and were interpreted qualitatively by the researcher. Most participants had positive attitudes towards listening to podcasts and engaged in class activities actively and have gotten less anxiety to participate in the discussion panels related to the files they have listened. They clearly expressed their attitudes toward podcasts as a useful educational technology.

**Procedure**

Prior to the study, the second version of a quick placement test, developed by Oxford University Press and University of Cambridge Local Examinations Syndicate (Syndicate, 2001), was administered to put the subjects in their own levels. Fifty four upper-intermediate female learners were selected among 118 students as the participants of the study. They were assigned into experimental ($N = 26$) and control ($N = 28$) groups. Two weeks prior to the study, the SRAQ administered as the pretest. As podcasting was the treatment of the study, the researcher had chosen 13 files among 23 podcast files as supplementary material to be used asynchronously prior to each session of the class. The podcast files were about personal development and they had been developed by Pavlita (2005). There was no need to listen to the podcast files online but it could be listened asynchronously. The students in the experimental group had the opportunity to choose the time they want to listen. Moreover, the students selected the order of the podcast files that the researcher assigned them to be heard during the term. In the control group the learners were asked to listen to a radio program in English for thirteen times before each session. All activities in both groups were completely the same the only difference was the type of treatment in the experimental group. Based on the teacher’s instruction there was no linear direction to do the activities and students were free to return to any part they needed and follow the process as they have required. The teacher did not teach the content of the podcasts and the students listened to the files prior to their class time. They were used as supplementary materials apart from their main course book. There was just a quick review and evaluation during the course. The key words were entered into the latest version of the SPSS software and were interpreted qualitatively by the researcher. Most participants had positive attitudes towards listening to podcasts and engaged in class activities actively and have gotten less anxiety to participate in the discussion panels related to the files they have listened. They clearly expressed their attitudes toward podcasts as a useful educational technology.
• The teacher suggested to the students to reflect on the topic and recall their feeling or personal experience toward that podcast and radio program and its content.
• The teacher asked the individuals to state their understanding of the topic and created a free discussion regarding to the podcast they had listened to.
• Students engaged in multi-task activities each session.

The researcher used 21 version of SPSS to analyze the quantitative data obtained of SRQ. Per session each learner recorded her own attitudes toward the use of podcast as an educational technology and wrote down her feeling about this type of learning. They also recorded their progress in learning the language and the changes happened in their learning strategies and behavior toward learning English language. At the end of the study all data were analyzed both quantitatively and qualitatively by the researcher.

**Results and discussions**

The data collected through instruments were analyzed by using SPSS Software to test the hypotheses of the study.

**Comparing the results of QPT in control and experimental groups before treatment**

In order to compare means of control and experimental groups in QPT before the treatment, independent samples \(t\)-test was conducted. Data are displayed in Table 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>(t)</th>
<th>Sig. (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>28</td>
<td>43.04</td>
<td>2.64</td>
<td>52</td>
<td>.94</td>
<td>.34</td>
</tr>
<tr>
<td>Experimental</td>
<td>26</td>
<td>43.73</td>
<td>2.73</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Table 1 shows, there is not any statistically significant difference (\(df = 52, t = .94, Sig. = .34 > .05\)) between control \((N = 28, M = 43.04, SD = 2.64)\) and experimental \((N = 26, M = 43.73, SD = 2.73)\) groups in QPT which confirms the homogeneity of the participants.

**Results of Self-Regulation Questionnaire (SRQ)**

In order to assess levels of self-regulation in pretest and posttest, paired and independent samples \(t\)-tests were conducted.

**Comparing results of SRQ in the experimental group before and after treatment**

In order to compare results of SRQ in the experimental group before and after treatment, paired samples \(t\)-test was conducted (Table 2).

<table>
<thead>
<tr>
<th>Pair</th>
<th>N</th>
<th>(M)</th>
<th>SD</th>
<th>df</th>
<th>(t)</th>
<th>Sig. (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>26</td>
<td>208.38</td>
<td>14.28</td>
<td>25</td>
<td>8.09</td>
<td>.000</td>
</tr>
<tr>
<td>Posttest</td>
<td>26</td>
<td>228.96</td>
<td>9.71</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Table 2 shows, participants in the experimental group \((N = 26)\) performed significantly \((df = 25, t = 8.09, Sig. = .000)\) better in posttest \((M = 228.96, SD = 9.71)\) compared to pretest \((M = 208.38, SD = 14.28)\) which confirms the significant effect of the treatment.
Comparing the results of SRQ in the control group before and after treatment

In order to compare results of SRQ in the control group before and after treatment, paired samples t-test was conducted (Table 3).

<table>
<thead>
<tr>
<th>Pair</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>t</th>
<th>Sig. (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>26</td>
<td>207.39</td>
<td>13.93</td>
<td>27</td>
<td>3.31</td>
<td>.003</td>
</tr>
<tr>
<td>Posttest</td>
<td>26</td>
<td>214.64</td>
<td>18.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Table 3 shows, participants in the control group (N = 28) performed significantly (df = 27, t = 3.31, Sig. = .003) better in posttest (M = 214.64, SD = 18.94) compared to pretest (M = 207.39, SD = 13.93). The t-value calculated through paired samples t-test for the experimental group (t = 8.09) was higher than that for control group (t = 3.31). Also, p-value was higher in control (p = .003) compared to experimental (p = .000). To compare means of control and experimental groups in SQR pretest and posttest independent samples t-test was conducted.

Comparing the results of SRQ in the control group and experimental groups before treatment

In order to compare levels of self-regulation of the two experimental and control groups before treatment independent samples t-test was conducted. Results are shown in Table 4.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>t</th>
<th>Sig. (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>28</td>
<td>207.39</td>
<td>13.93</td>
<td>52</td>
<td>.25</td>
<td>.79</td>
</tr>
<tr>
<td>Experimental</td>
<td>26</td>
<td>208.38</td>
<td>14.28</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Table 4 shows, there is not any statistically significant difference (df = 52, t = .25, Sig. = .79 > .05) between control (N = 28, M = 207.39, SD = 13.93) and experimental (N = 26, M = 208.38, SD = 14.28) groups on the SRQ pretest (before treatment) which confirms the homogeneity of the participants.

Comparing the results of SRQ in the control group and experimental groups after treatment

In order to compare means of control and experimental groups in SRQ posttest, independent samples t-test was conducted (Table 5).

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>t</th>
<th>Sig. (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>28</td>
<td>214.64</td>
<td>18.94</td>
<td>52</td>
<td>3.45</td>
<td>.001</td>
</tr>
<tr>
<td>Experimental</td>
<td>26</td>
<td>228.96</td>
<td>9.71</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Table 5 shows, participants in the experimental group (N = 26, M = 228.96, SD = 9.71) performed significantly (df = 52, t = 3.45, Sig. = .001 < .05) better than those in the control group (N = 28, M = 214.64, SD = 18.94) in the SRQ posttest which confirms the significant effects of implementation of the treatment throughout the course.

Self-reflection journals results

Regarding the second research question (learners’ attitude towards podcasting as an educational technology) participants in the experimental group were asked to write down their own voices in a self-reflection journal each session, including two reflection questions to express their own ideas about this experience. The researcher extracted the key words from the participants’ journals and elicited the participants’ views on the implementation of the treatment. To do so every individual’s attitude towards use of podcasts was categorized into three subfields as shown in Table 6.
Table 6. Results of self-reflection journals

| Learners attitudes towards Podcast as a tool (Educational Technology) | Not stated | | Stated | | Frequency | Valid percent | Frequency | Valid percent |
|---|---|---|---|---|---|---|---|
| Helpful/Useful | 2 | 16.7 | 10 | 83.3 | | | |
| User friendly | 1 | 8.3 | 11 | 91.7 | | | |
| Convenient | 2 | 16.7 | 10 | 83.3 | | | |
| Portable | - | - | 12 | 100 | | | |
| Facilitator | 4 | 33.3 | 8 | 66.7 | | | |
| Time shifting opportunities | 2 | 16.7 | 10 | 83.3 | | | |
| Support personalized learning | 4 | 33.3 | 8 | 66.7 | | | |
| Clear sound | 5 | 41.7 | 7 | 58.3 | | | |
| Free and economical | - | - | 12 | 100 | | | |
| Enrich computer skills | 2 | 16.7 | 10 | 83.3 | | | |
| Average percent | | | | 81.66 | | | |

| Podcast content | Not stated | | Stated | | Frequency | Valid percent | Frequency | Valid percent |
|---|---|---|---|---|---|---|---|
| Multi-tasking | 1 | 8.3 | 11 | 91.7 | | | |
| Interesting and enjoyable | - | - | 12 | 100 | | | |
| Tangible | - | - | 12 | 100 | | | |
| Motivating | - | - | 12 | 100 | | | |
| Stimulate practices | 4 | 33.3 | 8 | 66.7 | | | |
| General knowledge development | - | - | 12 | 100 | | | |
| Informative | 3 | 25 | 9 | 75 | | | |
| Authentic aural language | - | - | 12 | 100 | | | |
| Enhance discussion potential | 1 | 8.3 | 11 | 91.7 | | | |
| Average percent | | | | 91.67 | | | |

| Podcast application perception toward its effect on behavioral condition of the learners | Not stated | | Stated | | Frequency | Valid percent | Frequency | Valid percent |
|---|---|---|---|---|---|---|---|
| Personal development | - | - | 12 | 100 | | | |
| Self-awareness | - | - | 12 | 100 | | | |
| Humanistic point of view | - | - | 12 | 100 | | | |
| Less anxiety | - | - | 12 | 100 | | | |
| Individual learning | - | - | 12 | 100 | | | |
| Equal opportunity for shy and poor learners | 2 | 16.7 | 10 | 83.3 | | | |
| Change beliefs toward learning | - | - | 12 | 100 | | | |
| Collaborative/team learning | 1 | 8.3 | 11 | 91.7 | | | |
| Enhance self-confidence | 1 | 8.3 | 11 | 91.7 | | | |
| Average percent | | | | 96.3 | | | |

As Table 6 shows concerning the participants’ attitudes toward use of podcasts through their self-reflection journals learner’s attitudes has been classified into two main categories.

- Learners attitudes towards Podcast as a tool (Educational Technology)
- Podcast Content
- Podcast application perception toward its effect on behavioral condition of the learners

Concerning the first classification, in an average 81.66% of the participants written that podcast is a helpful and useful technological tool in learning a language which is user-friendly, convenient, portable, facilitator, free and economical with clear sound and could provide them time shifting opportunities and support personalized learning as well as enrich their computer skills.

Regarding the second class of learner’s attitudes, 91.67% believed that the content of the podcasts has a multi-tasking nature which seems interesting and enjoyable, tangible and close to their real life situation and experiences, motivating and stimulate more practices, informative and develop their general knowledge, provide authentic audio material and enhance their discussion potential.
With respect to the last classification of the learner’s attitudes towards podcasts a kind of personal development has occurred in themselves due to facing with ideas and humanistic viewpoints they listened and they got self-awareness about their own attitudes and beliefs. Since the atmosphere of the class has got less anxiety and individual learning happened to do the activities as well as equal opportunities for shy and poor learners they have changed their beliefs toward learning. Listening to podcasts engaged them in collaborative learning and accelerated their self-confidence. (See the following samples of the three learners’ final pages of their self-reflection diaries)

Samar:
The podcasts were very interesting; I loved them so much, because they’re useful, portable and easily accessible. Their content was informative, they developed my general knowledge and I could feel them in my life so tangible. It developed my pronunciation, vocabulary, speaking, and listening, too. It is really better than traditional methods of learning English, in this way I could learn English enjoyable and more, therefore it will be helpful for my future. Through the use of podcasts I wrote a summary in a special notebook. If I didn’t use them may be I didn’t have these useful information and experience like now. Also, these podcasts helped me to understand that which ways are better to reach my goals.

Behnaz:
The podcasts I listened were so informative for me because I learned a lot of things about how to live and be a successful person in the future. The podcasts were interesting for me now I learn how to have self-management and how to set my goals. The ways I listened to podcasts were convenient and easily accessible. I love to use technology in learning things and these podcasts were so great and they had a good technical quality. The abilities I achieved after listening to these podcasts were a lot to improve my listening skill, I have learned many new vocabularies, I developed my pronunciation and I speak a little bit faster and I learned many new idioms.

Mohadesheh:
I really like to listen to them and when I practiced them with myself, I enjoyed them very much. It was for me to listen to them two or three times, but when I listen to them every time and everywhere I believed that if I practice more I can listen and understand them better. I asked the bosses of the school or institutes to facilitate classes with the required technology. They helped me to know myself and try to reach my goals in a good way. Now I want to say an important usage of podcasts and it is improving listening.

Conclusion

The results of this study have been analyzed as follows. The results achieved by Brown et al. (1999) self-regulation ability questionnaire (SRAQ) approved that listening to podcasts significantly affected Iranian upper-intermediate EFL learners’ self-regulation ability. This finding confirms the results of an investigation done by Istanto (2011) exploring the use of podcasts ease the self-paced learning and provide a tool for test preparation as well as a beneficial platform to help slower learners, hence the results of this study represents that applying podcast affected the development of SRA of Iranian upper-intermediate EFL learners. However, many studies were done before to investigate the effect of listening to podcasts on EFL/ESL learners, they were mostly concentrated on oral skills of language learners and language learning proficiency like the studies done by Kargozi and Zarinkamar (2014); Hasan and Hoon (2013); Vandergrift (2006); Kan (2010); Ashraf, Noroozi, and Salami (2011) in which they found that listening to podcasts can significantly enhance language skills of the learners and enrich their oral skills as well as their vocabulary domain and grammar knowledge. But the prior studies were not focusing on strategies for better language learning or teaching. Hence this study went beyond the previous fields and empirically found the effectiveness of listening to podcasts as a supplementary material and an educational technology means to foster the learners’ self-regulation ability. Therefore, the major difference of this research with other investigations regarding the application of podcasts in language learning is the result that highlighted the effect of listening to podcasts on SRA of the language learners.

To triangulate the findings of this study the learners’ voice were extracted from their self-reflection journals and was added to the results as the quantitative part of the study. The findings revealed that employing technological devices in learning a language can enrich the essential parameters in self-regulation ability such as cognitive and
metacognitive knowledge in line with Vandergrift (2006) exploring that listening comprehension ability may depend on metacognitive knowledge of the learners. O’Bryan and Hegelheimer (2009) determined that learner’s listening comprehension processes and strategies and the effect of repetition could grow metacognitive awareness of the students over the course. The results of this study are in line with the finding of other scholars who tried to find the effect of using of other technological tools in language learning within a blended language learning environment such as mobiles. Shoa, Abdin, and Pour-Mohammadi (2011) found that mobile application can enhance learners’ motivation toward language learning by providing plenty opportunities for learning. Also Mohammadzadeh (2010) introduced podcasting as friendly technology that encourage learners to use its content during their language learning.

The participants pointed out detailed records of their feeling toward listening to podcasts within their self-reflection journals. They stated that every new session, they became more enthusiastic to involve in class activities with less anxiety. The participants had positive ideas toward the employment of technology in EFL classrooms. According to them technology enhances EFL students’ autonomy during the process of learning. Hence it is recommended that EFL teachers need to receive appropriate trainings on how to employ more technology in the classroom.

During the recent century use of technology has grown fast in many societies and going to be widespread in educational and learning settings. Podcast as a kind of technological tool can be applied in many gadgets to be used anywhere at any time and the learners in this study declared that they are keen to use podcasts, some of them listened to podcasts through their cell phones, some through their PCs, some on tablets, some on MP3 or MP4 players or other gadgets, at the bus, on the bed, in the car or wherever possible. They said that this freedom is so great to cope with their requirements and situation to enrich their learning process and develop their computer skills, too. These findings highlighted in agreement with Tayebinik and Puteh (2012) highlighted the use of various mobile devices in educational settings especially in TESL classes. Based on their review literature, it is clear that mobile learning has been adopted by language teachers as a beneficial tool to enhance language learning. Mobile phones are available and portable everywhere, hence this application is going to be widespread in every educational environment. A combination of English teaching and learning with mobile learning offers more creative and novel trends in language learning.

The results of this study showed that listening to the podcast is a new way of teaching and learning English language which can improve self-regulation ability of the learners as well as their computer skills implicitly. This implies that the use of technology can lead to higher self-regulation ability, learner and technology led atmosphere and facilitate the language learning process.

**Pedagogical implications**

There are some strategies used by individual successful learners referring to self-regulation ability. Zimmerman and Schunk (2011) and Kitsantas, Steen, and Huie (2009) findings revealed that the use of self-regulation strategies among primary school students could create considerable alterations in learners’ academic success. Therefore, by looking at the results of this study it is recommended to use materials, strategies, and approaches that could improve learners SRA to accelerate their learning process and become a successful learner.

The results of this study can be useful not only for EFL/ESL educational system/centers, but also for almost every educational system/centers and even for learners themselves who are developing their knowledge and skills by self. There are some explicit and implicit implications for designing the syllabus, producing materials, developing curriculum, teaching and learning design and processes. The outcomes can also increase teachers’ understanding of the effect of listening to podcasts on the development of EFL learners’ SRA in agreement with Arsal (2010) showing that the teachers play a significant role to improve learners’ self-regulated learning behaviors. Moreover, teachers can gain a vision about the impact of podcasts on improvement of self-regulation ability; other scholars can do various studies to investigate the methods that can help the learners regulate their behaviors to enhance learning.

This study suggests equipping EFL classrooms with computer and internet in order to have access to up-to-date authentic materials. Use of smart-board is another comment. Since the employment of technology in EFL classrooms is suggesting and it has been approved that technology enhances EFL students’ autonomy during the process of learning also EFL teachers need to receive appropriate trainings on how to employ technology in their classroom to
trigger some changes in traditional trends of learning and teaching and alter their trends into the use of appropriate educational technology. Educational settings ought to move from teacher centered classes into learner-centered. Through this approach, discovery learning would happen and both teachers and learners would be critical thinkers who reflect on their own activities, plans, and performance. The learners will set goals for themselves and find out the proper way on how to achieve their goals and the power would be shared among the teachers and learners. Innovation-ability would also foster through this approach which focused on shared decision-making through which instructors and learners negotiate all the class procedures in terms of content, learning strategies, evaluation process, tasks and activities. Moreover, this study directs the teachers to not restrict themselves to class activities and offer them asynchronous learning which can take place at any time or space. The teachers must pay attention to the learner’s opinions, feeling, interests, experiences, and abilities by offering them a chance to express themselves and eliciting learner’s idea to develop effective learning and boost the autonomy, self-confidence, self-regulation, as well as cognitive and meta-cognitive abilities of the learners.

Suggestions for further study

According to the findings of this study, some further studies are suggested to do as follows:

- Do other studies with both genders at various levels of language proficiency, different ages, and settings.
- Examine the effect of podcasting on different subject areas at school or university.
- Do other researches on other affective domains.
- Investigate the impact of podcasting on self-learning and lifelong learning process.

References


Curriculum Integration of MALL in L1/L2 Pedagogy: Perspectives on Research

Shu-Mei Gloria Chwo¹, Michael W. Marek² and Wen-Chi Vivian Wu²*  
¹HungKuang University, Taichung City, Taiwan // ²Wayne State College, Nebraska, USA // ³Providence University, Taichung, Taiwan // schwo@hk.edu.tw // mimarek1@wsc.edu // wcwu@pu.edu.tw  
*Corresponding author  
(Submitted May 22, 2015; Revised September 1, 2015; Accepted September 13, 2015)

ABSTRACT  
This study surveys work that has been done in the field of Mobile Assisted Language Learning. The researchers surveyed 70 corresponding authors of past MALL studies with formatted and open-ended questions, treating them as expert “participant-observers” of their own studies. The findings present details from the respondents about the MALL environments of the studies, the extent to which the MALL technology used in the studies has subsequently been incorporated into the curriculum, factors that positively or negatively affected such MALL integration, and insights of the respondents about the benefits of MALL. Although respondents cited a wide range of potential benefits from MALL use, most indicated that there was no actual curriculum integration resulting from their study, or that MALL was only integrated at the level of a single class meeting. This dismal level of integration suggests that researchers are using MALL for short-term tests and that researchers are exposing students to MALL to achieve academic publications, but with little intent to use it over the long term. The authors propose a framework of conditions, many or all of which must be met, in order to allow broad curriculum integration of MALL in the future.

Keywords  
MALL, CALL, Educational technology, Curriculum integration, Pedagogy, Affordance

Introduction

In the rapidly changing technological environment of developed countries, smartphones and other personal electronic devices are commonplace, particularly among young people. Many educators are moving to take advantage of the functionality of smart phone applications for Mobile Assisted Language Learning (MALL). MALL allows learning to occur conveniently and continuously in and out of the learner’s daily life and connect formal in-class lessons with informal learning outside the classroom (Wong et al., 2012). The continually changing field of smartphones and mobile applications, however, makes it hard for language learning researchers, let alone teachers, to remain current on the available technology and how it can be applied to teaching and learning. It is timely, therefore, to survey the work that has been done in this area, focusing on issues such as what mobile devices are actually being used, for what purposes, where, and what factors affect the feasibility of MALL implementation in connection with informal or formal language learning.

Leaders in the field of Computer Assisted Language Learning (CALL) have been calling for less experimental research into differences, comparing when a given technology is or is not used, and more research to determine affordances that a given technology can provide toward achieving outcome goals (Colpaert, 2012). The logical extension of this idea is that extensive meta-analysis is required to allow instructional designers to understand the affordances provided by various CALL systems, and best practices for their use (Marek & Wu, 2014).

The researchers concluded that a meta-analysis of MALL research, focused on identifying trends, actual contributions resulting from MALL research, and most recent practices, would be a valuable contribution to the MALL literature. The goal of the study, therefore, was to explore the perspectives of the corresponding authors of past peer-reviewed journal articles and conference papers with regard to the integration of MALL into the L1 and L2 classroom. The following specific research questions guided this study:

• What are the most common elements of MALL instructional design?
• To what extent have the outcomes of MALL research been incorporated into the curriculum?
• What factors have been found to positively or negatively relate to further MALL development and curriculum integration?
• Why do the respondents think MALL development and curriculum integration should be pursued?
Significance of the study

This study is significant because it surveys the corresponding authors of past studies who are arguably experts in MALL use, to identify patterns of MALL use as well as exploring the authors’ perceptions about MALL curriculum integration and factors influencing MALL development. The study is also significant because it reports not only technologies that have been tested, but the extent to which MALL systems are incorporated into sustainable technology design as opposed to being short-term “one-off” projects (Kennedy & Levy, 2009).

Finally, the study is significant because MALL research usually examines the opinions or perspectives of the students/participants. Few studies also explore the perceptions and reactions of the teachers/researchers, which is the primary data source of this study. Student perceptions may be skewed by convenience or easiness of learning activities, whereas the teachers and researchers were expected to be more focused on the learning goals and outcomes.

Literature review

In order to understand the findings and discussion of this study, it is necessary to understand the context of MALL as a learning technology. The authors, therefore, present a brief review of the benefits derived from using Computer Assisted Language Learning (CALL), benefits from using Mobile Assisted Language Learning (MALL), and a review of the typical MALL research context.

Benefits of CALL for EFL

Instructional designs which employ CALL have been shown to engage and motivate students, particularly when using the Internet to promote inter-cultural communication (Peterson, 2009; Wiebe & Kabata, 2010). Smartphones, G3/G4 connectivity, and ubiquitous WiFi have made 21st Century students continual users of mobile technology for their personal social interaction with friends and family. Research has shown that such Internet-based CALL applications can improve the skills of EFL learners (Terrell, 2011). It is natural to explore the use of evolving mobile technology for language learning, but the teachers and instructional designers often lag well behind the students in their sophistication in technology use.

MALL application use

The first smartphones, as we know them today, came on the market in 2001 (Conabree, 2001) from the Swedish company Ericsson, allowing text messaging and Internet connectivity. In 2003, 3-G cellular technology became available (Vodafone, 2013) providing relatively high bandwidth Internet connectivity for personal electronic devices.

Educators soon began exploring the capabilities of mobile devices. Research has found that learning activities using mobile devices allow students to be active participants in learning, not simply passive receivers of lecture content (Looi et al., 2010). The easy access provided by mobile devices allows the students to collect data, organize, and communicate synchronously and asynchronously, and access language resources conveniently (Wong & Looi, 2010).

Even though many studies have found merits in using MALL for learning, most MALL applications that have been tested were not created to serve as educational tools and adapting them to language learning may be demanding (Tai, 2012). In addition, when students are forced to undergo an extensive learning curve for a mobile technology which is new to them, it has been shown to reduce motivation (Looi et al., 2010).

Wong and Looi (2010) concluded that extending learning outside the classroom via MALL technology remains in its infancy. They predicted, however, that MALL has great potential to stimulate synergy by combining formal/in-class and informal/out-of-class learning in a way that transforms the language learning environment.
Context of MALL research

Even though many studies have confirmed the capability of mobile technology to enhance language learning, most studies have been conducted outside the classroom (Chen, 2014). In other words, many studies of MALL applications have not been integrated into classroom instruction or the curriculum (Viberg & Grönlund, 2013). Typically, volunteers are recruited or participate in the study by the researchers or the designers of the mobile programs and an external incentive reward is provided. This de-contextualizes the learning, decoupling it from the classroom, syllabus, and curriculum design, so there is often no meaningful long-term learning involved.

The above review of literature led the authors to conclude that a study exploring the insights of MALL researchers would make a valuable contribution to the scholarly literature. The following section describes the methodology used in this study.

Methodology

The researchers, working with other collaborators, had previously compiled a corpus of MALL studies used for EFL/ESL instruction from 1994 to 2012, collecting a total of 329 studies with 186 primary authors, including peer-review journal articles and conference papers (see Burston, 2012). The previous team used a variety of sources to identify journals, conferences, and specific papers, including keyword searches in multiple library and research databases, Google Scholar, and following reference links contained in previously-identified papers. An attempt was made to be as comprehensive as possible, such as including conference papers in the corpus in order to analyze as many diverse data sources as possible.

The current researchers made a concerted effort to locate the corresponding authors of the studies. Some were not located due to retirement or moving to different universities. The 138 MALL research paper corresponding authors who were located were asked to participate in a mixed quantitative and qualitative online survey using the SurveyMonkey platform. The survey was created based on a review of the relevant literature and the experience of the researchers. In order to establish validity, several external MALL experts reviewed the instrument, and certain changes were made as a result. The external reviewers were all published researchers in the field of computer assisted language learning, including one who was an associate editor of multiple CALL-related journals.

The twenty-eight questions covered the following categories: program details, subsequent curricular integration, factors influencing subsequent curricular integration, and general comments. Some questions asked for fill-in-the-blank information, such as the year of the study. Others used a Likert-like scale, ranging from strongly disagree to strongly agree. The final two questions were open ended. The complete survey protocol is in Table 1. A subset of questions, Q16-Q25, reflected possible factors influencing curriculum integration of MALL, proposed by the researcher based on the literature review and their own experiences.

Four of the completed surveys represented studies conducted in the 1990s, six were in 2000-2005, ten were in 2006, nine were in 2007, six in 2008, nine in 2009, nine in 2010, nine in 2011, and three in 2012 (although more studies from 2012 data may have been awaiting publication at the time of data collection for the current study).

After some prompting, seventy online survey responses were completed, for a response rate of 51%. The survey was administered in English and all responses were in English.

Descriptive statistics for the closed items were analyzed in SPSS. For research question one of this study, about the environment of the studies conducted by the participants, simple means were calculated. For research question two, analysis included p-value and Spearman’s rank correlation coefficient, in order to provide two different measures of correlation for the degree of MALL incorporation into curriculum design, and the factors which influenced MALL use. For research questions three and four, the researchers used the open-ended qualitative questions and performed content analysis, reading the responses several times and grouping key ideas into themes.
Table 1. Survey protocol

<table>
<thead>
<tr>
<th>Number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>You may report the results of this survey with identifying details.</td>
</tr>
<tr>
<td>2</td>
<td>Project/application name, if any.</td>
</tr>
<tr>
<td>3</td>
<td>Name of lead developer(s)/researcher(s).</td>
</tr>
<tr>
<td>4</td>
<td>Year the study began.</td>
</tr>
<tr>
<td>5</td>
<td>Year ended (or “ongoing”).</td>
</tr>
<tr>
<td>6</td>
<td>Mobile device (type).</td>
</tr>
<tr>
<td>7</td>
<td>Language being studied.</td>
</tr>
<tr>
<td>8</td>
<td>Language type (Native, second, foreign).</td>
</tr>
<tr>
<td>9</td>
<td>Proficiency level of targeted learners.</td>
</tr>
<tr>
<td>10</td>
<td>Primary project orientation (System design, lab trial, prototype, etc.).</td>
</tr>
<tr>
<td>11</td>
<td>Targeted learning environment (Primary, secondary, college, etc.).</td>
</tr>
<tr>
<td>12</td>
<td>Targeted learning area (Vocabulary, listening, writing, etc.).</td>
</tr>
<tr>
<td>13</td>
<td>Initial level of your project’s curricular integration.</td>
</tr>
<tr>
<td>14</td>
<td>We did/do not intend to integrate the mobile technology into any curriculum on completion of the study. (If no integration, end of the survey.)</td>
</tr>
<tr>
<td>15</td>
<td>The original project ultimately led to curricular integration at the following level (Entire program, entire course, individual class, none).</td>
</tr>
<tr>
<td>16</td>
<td>The pedagogical results of the initial project.</td>
</tr>
<tr>
<td>17</td>
<td>Reactions of students and faculty to the initial project.</td>
</tr>
<tr>
<td>18</td>
<td>The encouragement of administrative authorities.</td>
</tr>
<tr>
<td>19</td>
<td>The financial support obtained from the institution and/or commercial producers.</td>
</tr>
<tr>
<td>20</td>
<td>The technological infrastructure support obtained either from the institution and/or commercial producers.</td>
</tr>
<tr>
<td>21</td>
<td>Performance of relevant mobile hardware and/or software.</td>
</tr>
<tr>
<td>22</td>
<td>The willingness of colleagues to integrate mobile technology into the curriculum.</td>
</tr>
<tr>
<td>23</td>
<td>The pedagogical training of colleagues.</td>
</tr>
<tr>
<td>24</td>
<td>The technological expertise of colleagues.</td>
</tr>
<tr>
<td>25</td>
<td>The flexibility of the curriculum.</td>
</tr>
<tr>
<td>26</td>
<td>Factors which facilitated or inhibited further MALL development and curriculum integration.</td>
</tr>
<tr>
<td>27</td>
<td>Reasons to pursue MALL development and curriculum integration.</td>
</tr>
</tbody>
</table>

Findings

The research questions in this study were answered using a mixed methodology of quantitative and qualitative methodology. The data is presented below, divided by research question.

Research question one

To answer research question one, about the environment of the studies conducted by the respondents, simple means were calculated for questions which had been answered by Likert-like scales ranging from strongly disagree to strongly agree. It should be remembered that the environment findings span the full date range of the study and may not always reflect 2015 realities. In all cases where the means total more than 100%, it indicates mixed groups of students and/or multiple overlapping learning elements. The following section summarizes key points.

Smartphones overwhelmingly outnumbered other types of mobile devices and were used in 50.7% of the studies, followed by basic phones at 31.3%, and iPads or other tablet computers at 25.4%. Studies of foreign language learning constituted 72.1% of studies, followed by second language learning at 44.1%. Only 7.4% of the studies addressed MALL learning in the students’ native language. Many languages were studied, dominated by English 77.3%, followed by French 13.6% and Chinese 12.1%. Several other languages were in single digits, including Basque, German, Greenlandic, Hmong, Italian, Japanese, Spanish, and even American Sign Language.

Figure 1 shows the breakdown of categories of outcome goals, including the four basic skills of English and other targeted learning areas. Vocabulary constituted two-thirds, 66.7%, followed by listening at 55.1%, speaking at
37.7%, reading 36.2%, and grammar, 30.4%. Culture was a learning focus for 29.0% of the studies and writing 24.6%. As noted above, in many cases more than one area was targeted, resulting in totals higher than 100%.

Most of the studies included participants of intermediate proficiency, accounting for 69.6% of the studies. Participants were beginners in 53.6% of the studies whereas 15.9% of students had advanced proficiency. Again, many studies included participants of more than one proficiency level, resulting in a total higher than 100%. College students made up 69.1% of the participants, followed by 10.3% in primary school, and 8.8% each in adult education and independent study. Only 2.9% were in secondary school.

The study examined several external factors influencing the MALL research projects. Of the respondents, 51.6% were neutral about the level of encouragement they had received from administrative authorities, 41.9% were either positive or strongly positive, while 6.4% were negative or strongly negative (Figure 2). Concerning the amount of financial support they had received, 43.6% of the respondents were either positive or strongly positive, 29.0% were neutral, and 17.5% were negative or strongly negative (Figure 3). Similarly, 42.6% were neutral about the amount of technical support they had received, 42.7% were positive or very positive, and 14.8% were negative or strongly negative (Figure 4). Two thirds, 67.2%, of the respondents were positive about the performance of the hardware or
software they had used, with an additional 13.1% strongly positive. Only 14.8% were neutral and 4.9% negative or strongly negative about their hardware and software (Figure 5).

![Figure 4. Technical support received](image)

![Figure 5. Hardware/Software performance](image)

The pedagogical results of the studies were reported to be 25.8% strongly positive, 62.9% positive, and 11.3% neutral (Figure 6). None of the participants indicated negative or strongly negative pedagogical results. Similarly, reaction of students and faculty were reported to be 33.9% strongly positive, 61.3% positive, 3.2% neutral, and 1.6% negative (Figure 7). In these two cases, the current researchers caution that the strongly positive trends may result from the fact that poor or negative results are rarely published in peer-reviewed journals.

![Figure 6. Pedagogical results](image)

![Figure 7. Reaction of students and faculty](image)
Research question two

Research Question 2, about the extent to which outcomes of MALL research have been incorporated into the curriculum, was answered by both quantitative and qualitative data.

Twenty-one-point-four (21.4%) of the respondents said that when they began their project, it was integrated into an entire academic program, 14.3% said it was integrated into an entire course, 22.9% said it was integrated into an individual class meeting, and 41.4% said it was not integrated at any level, meaning it was a short term test outside the curriculum.

The respondents reported that in 40.6% of the cases, there was never an intent to bring about integration (Figure 8). The respondents said that the technology in their study was intended to integrate MALL into individual meetings of classes in 23.2% of their answers, into an entire course 14.5%, and into an entire program for 21.7% of the studies. In response to the question about the nature of any curricular integration that actually took place, 48.1% indicated that none resulted from their research, 29.8% said there was integration during an individual classroom meeting only, 7.4% reported the technology integrated into an entire course, and 14.8% resulted in integration at the programmatic level (Figure 9).

![Figure 8. Intent to integrate after study](image)

![Figure 9. Actual integration accomplished](image)

Table 2. Correlation with MALL curriculum integration (Question 15)

<table>
<thead>
<tr>
<th>Question</th>
<th>Native language (n = 5)</th>
<th>Second language (n = 29)</th>
<th>Foreign language (n = 43)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rho</td>
<td>p</td>
<td>rho</td>
</tr>
<tr>
<td>16. Pedagogical results</td>
<td>.592</td>
<td>.293</td>
<td>.133</td>
</tr>
<tr>
<td>17. Reactions of students and faculty</td>
<td>-.148</td>
<td>.812</td>
<td>.219</td>
</tr>
<tr>
<td>18. Encouragement from the administration</td>
<td>.803</td>
<td>.102</td>
<td>.320</td>
</tr>
<tr>
<td>19. Financial support obtained</td>
<td>.811</td>
<td>.096</td>
<td>.300</td>
</tr>
<tr>
<td>20. Technological support</td>
<td>.725</td>
<td>.165</td>
<td>.175</td>
</tr>
<tr>
<td>21. Performance of software/hardware</td>
<td>.803</td>
<td>.102</td>
<td>.005</td>
</tr>
<tr>
<td>22. Willingness of colleagues to integrate MALL into curriculum</td>
<td>.632</td>
<td>.368</td>
<td>.216</td>
</tr>
<tr>
<td>23. Pedagogical training of colleagues</td>
<td>.316</td>
<td>.684</td>
<td>.207</td>
</tr>
<tr>
<td>24. Technological expertise of colleagues</td>
<td>.592</td>
<td>.293</td>
<td>-.231</td>
</tr>
<tr>
<td>25. Flexibility of the curriculum</td>
<td>.684</td>
<td>.203</td>
<td>.531</td>
</tr>
</tbody>
</table>

*Note.* *p* < .01.
To further answer research question 2, Spearman correlations were calculated to relate survey question 15, about whether the original project ultimately led to curricular integration, with survey questions 16-25, about possible factors influencing curricular integration of MALL. The correlations are shown in Table 2.

Based on the $p$-value, the most prominent factor favoring greater subsequent curriculum integration in studies using MALL for foreign language acquisition was the reaction of students and faculty ($p = .006$). Also notable in foreign language instruction, but with higher $p$-values, were encouragement from administration ($p = .057$), financial support ($p = .062$), performance of hardware/software ($p = .077$), and willingness of colleagues to incorporate MALL into the curriculum ($p = .087$).

In studies using MALL for second language acquisition, i.e. where there is a surrounding population of native speakers outside the classroom, flexibility of the curriculum stood out as the single crucial factor ($p = .004$). Almost all other considered factors had weaker but positive correlations with curricular integration.

In MALL use for native language instruction, i.e. where there is a surrounding population of native speakers outside the classroom, flexibility of the curriculum stood out as the single crucial factor ($p = .004$). Almost all other considered factors had weaker but positive correlations with curricular integration.

**Research question three**

Research Question 3, about factors that had been found to relate positively or negatively to further MALL development and curriculum integration, was answered by qualitative analysis of the open ended answers to question 26. In the results below, the number in parentheses, e.g., (20), identifies the survey respondent making the comment, anonymized for privacy.

Respondents to the survey described a rich range of factors promoting or inhibiting MALL integration, resulting in a typology identified by the researchers as factors related to the institution, program, teacher, student, task, technology, and outcomes. Many considerations were mentioned positively, when they were in place favorably, as well as negatively, where they were absent or not working properly.

**Institutional factors**

Institutional factors were prominent in the comments about facilitation or inhibition of MALL programs. The most mentioned area of institutional responsibility was financial support. One respondent was “inhibited by total lack of meaningful financial support for my efforts” (3), but another observed that “three learners had their own personal iPads…In addition we received a University research grant covering the purchase of five iPads, accessories, and applications which were loaned to volunteer participants” (18). Though respondents are not always clear, hardware appears to have been the main required expense, suggesting that despite the rapid worldwide growth in sales of electronic devices, many projects rely on devices which not all the learners owned (e.g., smartphones, iPads). On the other hand, other costs were also mentioned, such as that of messaging (67), and buying online material (55).

Institutions have an influence in other ways as well, including through their general understanding of, and policies concerning, MALL. Respondents also highlighted the need for adequate infrastructure, such as WiFi, robust bandwidth, and technical staff members who are competent to assist with the MALL applications. Many of these factors could also be considered to be technical factors, but are included here because they are based on institutional budgets, decisions, and policies.

**Program-related factors**

The nature and flexibility of the curriculum, and other factors related to the academic program, is of paramount importance to whether a given teacher is able to innovate with MALL. As reported under the Institutional Factors
section, respondents most often cited financial support as positively or negatively influencing MALL integration. In most cases, the respondents did not specify whether this needed support was situated at the academic department or at the institutional level.

Respondents also listed the amount of time, or release time, available for actual development of a MALL course or curriculum as important. Furthermore, they reported that the nature of assessment is very relevant, including both the evaluation of the work of the teachers and the evaluation of students for the purpose of grades. Individual respondents mentioned negative factors including a programmatic culture of examination-oriented classes and teacher-centered learning (22), a program driven by technologists and not by language instructors (26), and a lack of consistency across classes with the same subject name but taught by different teachers (51). Others saw little flexibility in the overall curriculum (40) or that “no one cared what I did in the classroom” (45). One bluntly said “there was never an intention to change the curriculum” (18).

Teacher-related factors

The interest/enthusiasm of the teacher was often mentioned by the respondents, usually coupled with recognition that this motivation may not be shared by all teachers, making teacher permanency an important factor. One teacher noted that, “In our experience, curricular integration of MALL is linked to the willingness of individuals…Long term integration might not be foreseeable when these individuals are not permanently attached to the institution” (19). Most comments about teacher-related factors supported MALL, but one respondent voiced a lack of enthusiasm, saying “Teachers fear additional work with mobile technology” (63).

Aside from the teachers’ acceptance, their “epistemological belief” (44) was also mentioned as being crucial. MALL is based on a theoretical framework of Constructivism and student-centered active learning, with considerable student autonomy, which the respondents said the teacher needs to embrace in order to achieve success. Furthermore, the respondents called attention to the need of the teacher for specific requisite knowledge such as Technological Pedagogical Content Knowledge (Harris, Mishra & Koehler, 2009), i.e., an understanding of how technology, pedagogy, and content can inter-relate. Teacher 41 remarked that, “(1) Language teachers are left behind the IT technologies. (2) Business sectors are much more enthusiastic to develop a new style of language learning, which is still being developed.” Another respondent highlighted the resulting need for professional development with MALL systems for part-time staff (69).

Student-related factors

Students are also crucial to MALL integration and their interest/motivation was of importance to the respondents. The respondents, however, did not always find student factors to be positive influences on integration, due to the complexities of student motivation and engagement. The respondents found that the students often had heavy workloads that reduced their motivation for use of a new technology (54), and that in years past, some students did not own their own smart devices (58). The respondents recognized that student culture often treated use of mobile technology as a “fashion trend” (23), which could be a positive indicator for MALL. However, the respondents also found that the “Students’ cultural dependency on mobile devices both facilitated and inhibited…Facilitated because [students were] free to use devices for formal learning…Inhibited if the level of privacy intrusion in usage was high” (12).

This point was further illuminated by another respondent who said, “some students do not want to receive SMSs related to study, and/or do not want a “push” use of technology but prefer to “pull,” i.e., choose for themselves when to access material…Students’ reactions to the pushing of messages were too diversified for us to pursue it further; lots found it very useful and stimulating but quite a few did not take to it at all” (67). In other words, the very fact that students are already using various devices such as mobile phones in their everyday life, typically for social and leisure purposes, means that they have an established set of expectations or unwritten rules which are followed in their use. For some this means that they expect primarily to initiate interactions (i.e., pull content) and they do not expect to be the object of material imposed on them (i.e., pushed to them), which they see as an invasion of privacy, similar to SPAM email.
As with the teachers, students not only need interest but also a positive frame of mind to accept instructional designs using MALL. Respondent 44 pointed to the need for a certain “epistemological belief” on the part of the student, and respondent 27 indicated that a reason for failure was that “students needed more exposure to non-didactic teaching approaches for our informal student-centered approach to be more successful.”

**Task-related factors**

There was little mention of the nature of learning tasks being relevant to successful integration of MALL technology in the curriculum. One respondent noted that “data also showed that [the iPad’s] usefulness depended on the task” (4) and another mentioned that it is important to have *specific* tasks for students to accomplish via MALL outside the classroom (45).

**Technology-related factors**

Many respondents identified Internet bandwidth as a technology-related factor that was important to MALL integration. Remaining comments about technology divided fairly equally between device issues (i.e., hardware/firmware) and software functionality. For the hardware, availability was satisfied if the students already owned the required devices. In some projects, however, this was not the case, especially projects requiring capability beyond those of basic phones, making the availability of needed features on the students’ phones an inhibiting factor (15). One study used an application that was only available for Android phones, but many students owned iPhones, or used other non-Android operating systems (16). Some respondents cited use of study-provided iPads to alleviate this problem (44). Other respondents cited a lack of software developers for customized applications (32); that a student developer graduated and the project became orphaned (21); Apple’s unexpected changes of XCode (31); and felt that technology, generally, moves too fast for teachers to keep up (35).

**Outcomes as a factor**

Finally, the respondents occasionally mentioned the outcome benefits of MALL use as a factor affecting integration. If students perform well in assessments, “positive outcomes on semester-end assessments with courses using MALL” (2) can influence implementation. Sometimes it was the outcomes experience of a specific technology, like the iPad offering speed, video viewing, and versatility, which influenced implementation.

The findings, therefore, show a diverse range of factors affecting the likelihood of integration of MALL technology into the curriculum. Many are specific to the environments of individual schools. While they cannot be considered to be generalizable, the qualitative data affirmed the typology of factors proposed by the researchers as influencing whether MALL is integrated into the class, course, or program-level curriculum. The typology, therefore, includes institutional factors, program-related factors, teacher-related factors, student-related factors, task-related factors, technology-related factors, and factors related to the outcomes when using MALL.

**Research question four**

Research Question 4, about reasons why the respondents feel MALL development and curriculum integration should be pursued, was answered by qualitative analysis of the open ended answers to question 26. In this question, the respondents were asked to comment in general about curriculum integration of MALL, not limiting their comments to their specific study or their specific institutions.

Some respondents gave general comments, stating that MALL would have positive effects (10). Others were more specific, saying that it would enhance learning (13, 19, 23); be generally useful/effective (25, 43, 50, 61); ease the learning process (9, 10, 54), even for learners who are not proficient with technology (62); that MALL integration would be in the best interests of the students (18, 19, 26, 1, 36, 41, 43, 47, 55), and in the teacher’s interest (26, 31, 66).
Another theme was that trends in everyday life make MALL integration almost inevitable in the educational sphere. One respondent said “It is what learners do!” (32), making the point that not integrating MALL would, in effect, be missing out on exploiting ubiquitous devices. Other respondents said that “young people like our students are sensitive to any new media technology, and they can easily introduce it to their daily lives; the educational environment is not an exception!” (38); and that “MALL is an integral part of young learners’ lives” (39). The flexibility of MALL was noted by the respondents (7, 68), specifically that it can be used everywhere (9, 56, 68), especially outside of class (29, 63). This means that free time is better utilized for learning than before (11, 12, 68). Consequently the respondents felt that time and effort were saved (23), as well as money (36) and even paper (56). One respondent indicated that there is a logical progression from oral storytelling to books to ball point pens to mobile devices (39). Another respondent said that the key incentive to implement MALL is when it eliminates problems with other instructional methodologies (51).

Many of the respondents cited specific areas of learning in which they feel MALL can provide benefits. They include language (6, 19, 22) and writing (17), but also more broad cultural understanding (6), EFL digital literacy (3), and Information and Communication Technology skills (ICT) (17). The respondents often saw MALL as best used for practicing the skills taught in class (45, 57). In other words, the respondents tended to present MALL as a platform for task-based additional/supplementary use of language, rather than a vehicle for the initial introduction to new material, which they characterized as best done in the classroom (25, 29, 36, 55).

Other respondents highlighted the benefits of MALL to enhance the authentic use of language (6, 25, 33, 55) and fitting in with activities learners do anyway (32, 47), especially for young learners (38, 39). Other adjectives used to describe MALL integration were that it is “contextual” (33), “task-based” (46), “interactive” (33), and “interdisciplinary” (46).

Some responses aligned with cutting-edge scholarly perspectives on foreign language teaching, as a whole, saying “The roles of teachers and students should be transformed” not just by the devices but the ways of using them (13) and one should “make the most of opportunities to bring together informal and formal learning” (57). Respondent 44 referred to “the need of curriculum reform in language learning (moving away from pure classroom-based instruction)...The need of blending language learning into daily life.” This is closely linked to the view that MALL integration enhances learner autonomy and self-directed learning (27, 39, 59, 63) and even that MALL can allow “students to be coauthors and co-creators of content” (46) as opposed to learning from materials that flow in one direction only, from teacher to student.

The respondents felt that it is often the high-level language learner who can benefit most from MALL integration. One said, “There’s a need for informal mobile tools at this high level of language learning, where learner needs may be quite specific and difficult to predict” (27, echoed by 39, 63). But respondents also notes other socio-economic sectors which could benefit from MALL, such as non-school-going immigrant learners (36) or students in remote areas who cannot attend classes (62).

This latter need was highlighted by respondent 36, who said, “Mobile learning seems to be a very relevant tool for reaching the typically not so ICT-literate in the very small fishing and hunting societies in Greenland with courses in Danish and English (targeted for tourism, work in mines, etc.).” The respondent described a plan in which a traveling teacher, with tablet computers and especially developed lessons and instructional material, would visit remote communities that have no computers or Internet.

The findings, therefore, show that respondents cited generally-accepted benefits of Computer-Mediated Communication (CMC) and Computer Assisted Language Learning (CALL), that MALL integration follows trends and expectations in society, and can provide authentic contextualized learning. Their responses emphasized the flexibility of MALL to address a wide range of instructional goals and outcomes.

**Discussion**

This study drew on the experiences and perceptions of researchers in past Mobile Assisted Language Learning (MALL) studies. In their original studies, they served as independent researchers, however in the current study they are “participant-observers.” The authors are aware of no other study that examines this perspective.
**Environment**

The answers to research question one, about the environment of the studies, are not surprising. Although technology patterns have changed over the years, nearly half of the respondents had published in the previous five years, making the overall data weighted toward current practice. Accordingly, it is no surprise that smartphones were the dominant MALL technology, followed by basic phones and tablet computers. The robust nature of English teaching around the world also makes it no surprise that well over two-thirds of the studies included foreign language learning, in which students have little opportunity outside the classroom to practice their language skills (Wu, 2006). The fact that two-thirds of the studies included vocabulary, and over half included listening, may reflect the ease with which such lessons can be incorporated into a MALL application. Speaking, reading, grammar, and cultural understanding are arguably more difficult to present and evaluate as MALL learning activities.

The most notable finding in the answer to research question one was that only about 5% of studies reported negative outcomes of the MALL technology used and only 1.6% of studies reported negative student or faculty opinions about the MALL technology. It is notable that none of the participants indicated negative or strongly negative pedagogical results.

The authors believe that this lack of negative outcomes is likely not the result of overwhelmingly positive results, but rather the result of researchers declining to report negative results, as well as peer reviewers who are skeptical about manuscripts reporting poor outcomes. The authors see this absence of negative outcomes as a failure of the EFL scholarship and peer review system, because scholars can often learn better from failures, or only partial success, than they can from unqualified achievement (Marek, 2014; Marek & Wu, 2014). A study may be well-designed and theory based, but still produce outcomes that are not completely satisfactory. In part, this is because there are innumerable factors operating outside the control of researchers, making them difficult to account for fully in a research design (Marek & Wu, 2014). Scrutinizing problems in a given instructional or research design, combined with critical thinking, can produce best practices that can be incorporated into future instructional technology design and research practices. Scrutinizing studies in which all factors seem to be ideal does not produce the same opportunities for improvement.

**Integration**

The data about integration of MALL technology into the curriculum shows that MALL is far from a mature field. Indeed, nearly four-fifths (77.9%) of respondents indicated that there was no integration at all, or that MALL was only integrated at the single class meeting level. This dismal level of integration suggests that researchers are using MALL for short-term tests in which researchers are exposing students to MALL with little intent to use it in the long term, i.e., they are conducting “one-off” studies disparaged by Kennedy and Levy (2009, p. 446).

The correlation analysis, meanwhile, showed that the reaction of students and faculty had the highest correlation with MALL integration, meaning that the better the students and faculty liked the MALL implementation in the study, the more likely that the faculty member would use it again, or that it would become a formal part of the academic program. Although the current study did not specifically ask respondents about their perception of the effectiveness of their MALL technology design, the authors conclude that when students and faculty both have positive experiences with a MALL technology, the technology design can be deemed to be effective, increasing its likelihood of being used again in the future. These findings suggest a high level of unfulfilled potential for MALL integration.

**Factors predicting integration**

The findings for research question three, on the other hand, demonstrate the complexity of achieving beneficial MALL integration. The typology of institutional, program, teacher, student, task, technology, and outcome-related factors includes a bewildering array of potential influences, positive and negative. As Marek and Wu (2014) pointed out, many internal and external factors operate on both the individual student and on the program/school, meaning that success can be hard to predict. Similarly, advocates of using sustainable MALL instructional design (Kennedy & Levy, 2009) in a course or program must negotiate a diverse array of factors in order to achieve success, not all of which may be easy to predict.
Participant-observer perceptions of benefits

Although academia has been slow to incorporate MALL into courses and/or programs, the participant-observers contributing their ideas to this study see many potential benefits from MALL incorporation. These include the benefits most frequently attributed to CMC and CALL (Bax, 2003; Sengupta, 2001) as well as an assortment of specific benefits that could flow from specialized application design. But the specialized ability necessary to design such highly customized MALL systems is often not within the skillset of faculty, and schools are often not able to mobilize the level of support, particularly when a new application is requested for use below the program level. On the other hand, custom applications may have learning curves which, if too steep, may demotivate students (Yu, 2011).

The problematic nature of crafting custom applications has caused scholars to explore a range of existing platforms not specifically designed for education, but which are widely available, such as Facebook (which is available in both MALL and non-MALL form) and other mobile platforms that can be adapted to learning activities. The result, however, evokes Colpaert (2012) and Dörnyei (2014), both of whom decry selecting the technology first and then trying to find learning activities to perform with it. These questions, again, show that there are many roadblocks to advanced curriculum integration of MALL technology, reinforcing the unfulfilled potential of MALL.

Required MALL framework

The overall findings of this study may seem discouraging, given the low levels of actual curriculum integration, even though the study shows broad potential for future MALL application use in language learning. Based on the findings of this study, and insights flowing from the discussion, the authors propose the following framework of conditions, many or all of which will need to be met in order for there to be broad curriculum integration of MALL in the future.

- Curricula philosophies must embrace task-based student-centered active learning in a Communicative Language Theory (Savignon & Wang, 2003) environment, with strong scaffolding. These are mainstream approaches to education in the 21st century, but are often not the instructional strategies used around the world in language learning.
- Professional development for faculty and administrators must show how MALL can benefit student learning, and provide practical strategies for development of lesson plans.
- Strategic curriculum design must be based on required student outcomes, with sustainable technology selected to provide affordances that address learning objectives.
- Administrative and budget support for technology choices must be determined by strategic planning.
- Technical support must be available from schools for application development and maintenance, including full software documentation. This area is particularly important because where such support is currently available it is often assigned to student programmers, resulting in lack of continuity when the student programmers graduate.
- Promotion and tenure policies must value curriculum design and encourage sustained use of MALL systems.
- Privacy questions stemming from use of MALL must be addressed, including potential visibility of student names and/or work by people outside the class, and the acceptability of classwork being stored off campus on servers of social media sites and application developers.
- Learning Management Systems (LMSs) should replicate common smart phone application functionality and ease of use, while remaining secure on the campus servers. This condition is not directly indicated by the data in this study, but many studies report that students are demotivated by complex learning curves of educational technology. LMS interfaces are often clunky and burdensome to use compared to popular apps and social media platforms.
- Academic and peer review standards must not marginalize studies that are well designed but not fully successful. As pointed out in the discussion, educators wishing to use MALL cannot fully understand MALL environments if the academic literature does not include discussion of problems and unexpected outcomes.
Conclusion

This study has presented a first-ever view of Mobile Assisted Language Learning from the perspective of researchers and educators who have taught using MALL. Their participant-observer expertise has revealed abundant opportunity for future development and integration of MALL, but also many impediments that often serve to block MALL from moving beyond short term “one-off” tests that are more stunts than valuable learning experiences for students. The authors hope that the reality check provided by these MALL participant-observers will open the door to discussion of changes in the academic environment that can pave the way for expansion of MALL as a mature educational strategy.

Acknowledgements

The authors appreciate the contribution of Dr. Jack Burston to this paper, including providing the SurveyMonkey account used to collect data, and his collaboration until January 2013; and the contribution of Dr. Glenn Stockwell, for his collaboration until June 2012.

References


Viberg, O., & Grönlund, Å. (2013, October). Integrating mLearning language application into university course content. Paper presented at the 12th World Conference on Mobile and Contextual Learning (mLearn 2013), Doha, Qatar.


Mining Learning and Crafting Scientific Experiments: A Literature Review on the Use of Minecraft in Education and Research

Steve Nebel*, Sascha Schneider and Günter Daniel Rey  
Department of E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Germany  
steve.nebel@phil.tu-chemnitz.de // sascha.schneider@phil.tu-chemnitz.de // Guenter-Daniel.Rey@phil.tu-chemnitz.de  
*Corresponding author  
(Submitted May 11, 2015; Revised July 29, 2015; Accepted September 17, 2015)

ABSTRACT  
Since the field of educational videogames or serious games is not limited to games that are specifically designed for educational purposes, videogames such as Minecraft have aroused the attention of teachers and researchers alike. To gain insights into the applicability of Minecraft, we reviewed the literature on use of the game in education and experimental research. We summarized the current usage in addition to our own considerable experience with Minecraft in courses on educational videogame design and as a research instrument in instructional psychology and discuss the benefits and limitations. Based on these observations, we outlined the future of Minecraft in both fields and emphasize examples that already stretch the technical and methodical boundaries. To increase the application of our analysis, we distill three main implications from our observations that address the future of educational and research tools in educational videogames in general.

Keywords  
Minecraft, Educational videogames, Serious games, Educational research, Educational technology

Introduction  
When analyzing educational videogames, researchers can observe that games, which cannot be described as designed specifically for educational purposes (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Ritzhaupt, Gunter, & Jones, 2010), have been used. Off-the-shelf games like Civilization III (Squire, 2004; Squire, DeVane, & Durga, 2008), Making History (Watson, Mong, & Harris, 2011), Massively Multiplayer Online-Role-Playing Games like World of Warcraft or EVE Online (Clark, Nelson, Sengupta, & D’Angelo, 2009; Rausch, Faschauer, & Martens, 2012; Reeves, Malone, & O Driscoll, 2008), simulation-based videogames like Sim City 2000 (Tanes & Cemalcilar, 2010), and narrative-focused adventures like The Walking Dead (Staaby, 2014) have been used in education or educational research. Even software that was never intended to be used as a game at all (e.g., Power Point, Sikó & Barbour, 2013; Sikó, Barbour, & Toker, 2011) appears in this context. These examples show the growing interest of researchers and teachers in using very different software as educational tools. Thus, we conclude that educational videogames are not only games that have been intended to be educational tools in the first place but also games that can be utilized as such. This can be elucidated even further when observing the impressive success of Quizkampen (http://www.quizkampen.se/), a mobile-based trivia game that includes online duels between friends and strangers. The players have to answer questions that address very different topics such as politics, geography, movies, or film stars. Thus, a stable learning process within topics that are not entirely entertainment-based is needed to improve in the competitive scenario.

The German version Quizduell has been downloaded more than 10 million times (SpotOn, 2014) and was extended to other media since the game is the center of a nationwide TV show (Das Erste, 2015). The success of a game that combines educational elements and competitive videogame mechanics is an appropriate example of the potential of videogames with entertainment-focused intent. Another videogame, Minecraft (https://minecraft.net/), had a significant impact not only on videogames in general but also on education and research. More than 19 million copies (Mojang, 2015a) have been sold for PC, 12 million copies for XBOX360 (4J Studios, 2014), and more than 21 million copies for mobile phones (Bergensten, 2014), which places the game on the all-time best-sellers list. Since its developer Mojang has been taken over by Microsoft for $2.5 billion (Owen, 2014), the game is known to the public as this was addressed in news around the world. In addition, millions of hours of gameplay footage and let’s plays are available for interested users who are either too young or do not have the technical or financial capacities to play the game themselves.
What is Minecraft?

In 2009, independent developer Zachary Barth created Infiniminer (http://www.zachtronics.com/infiniminer/) in his spare time. Within this game, he invented a new way of representing the game world with simple independent, block-shaped entities and added a procedural generation of the environment. Users who enter this randomly generated game world mine and place blocks, and thus recreate anything that can be constructed within this block-based structure. This process can be compared to a square pixel and low-resolution pictures. Based on a simple element, like a pixel, one can create almost every image he or she wants. Although the game includes player versus player gameplay, the simple but ingenious building mechanics and the world creation method inspired Markus “Notch” Persson to create his own Java-based version that eventually became Minecraft. This multiplayer sandbox building game (Ekaputra, Lim, & Eng, 2013) is simply “a game about braking and placing blocks” (Mojang, 2015a), but the game’s complexity has drastically increased since then (for a more extensive description see: Duncan, 2011).

More functionality through a diverse range of different block types was included, for example, the material of redstone transport signals, dirt blocks can grow grass, or coal can be used in ovens. Thus, players not only place blocks to build static representations of real-world elements but also use the blocks’ functionality to create complex machines or environments. Using a pixel-picture analogy again, the users not only build pictures but also add functionality, like animations. Additionally, server-based multiplayer was included so large numbers of players work together (or against each other) to create even bigger environments. Finally, survival mode and non-player characters (NPCs: both hostile and friendly) are part of the game. Although this game mode is not important for a large part of the community (who prefer the creative mode), the mode serves as an important starting point for the players’ experience within the game and, therefore, for its success. When the players start in survival mode, they have to build a simple shelter in order to stay alive. After creating their first small building, they are tempted to improve it even further. Thus, survival mode infuses user creation, which, in turn, develops a momentum of its own and becomes an essential part of the experience (Quiring, 2015). This momentum, the modification friendliness of the simple Java application, and the creative mode, which enables unlimited resources, have led to a tremendous amount of user-created content and aided Minecraft’s success. Starting with automatic factories that were used to make the gameplay tasks easier (e.g., automatic crop harvesting, traps for monsters), players soon expanded the game’s boundaries. Projects such as a functional 32-bit calculator (Hendrix, 2014, Figure 1), the recreation of King’s Landing (cyborgratchet, 2013, Figure 2) from the HBO TV show and the book series A Game of Thrones (Benioff & Weiss, 2011; Martin, 1996) and an art platform like OPERAcraft (Bukvic, Cahoon, Wyatt, Cowden, & Dredger, 2014) demonstrate the creative potential of the game and the effort some users put into their creations.

Figure 1. The display and parts of the necessary logic circuits of a 32 bit calculator
How Minecraft is used in education

Mojang is not the first videogame developer to enter the educational field with their products. Other developers like Valve or Electronic Arts started projects like STEAM for SCHOOLS (http://www.teachwithportals.com/) or supported modified versions of their games like SimCityEDU (https://www.glasslabgames.org/games/SC) to use the educational potential of their entertainment-based content. In contrast to these cautious advances, Minecraft is already in use as an educational tool for very different topics all over the world (Minecraft Teachers, 2015; Short, 2012). Minecraft has been used to enable early access to the topic of spatial geometry during class level 5/6 (Fürster, 2012), to teach about sustainable planning (West & Bleiberg, 2013), language and literacy (Bebbrington, 2014; García Martínez, 2014; Hanghøj, Hautopp, Jessen, & Denning, 2014), digital storytelling (García Martínez, 2014), social skills (Petrov, 2014), informatics (Wagner, 2014), computer art application (Garcia Martinez, 2014), project management (Saito, Takebayashi, & Yamaura, 2014), and chemistry (Hancl, 2013).

Further topics teachers want to address are ecology, geology (Ekaputra et al., 2013), biology, physics, geography (Short, 2012), arts, history, and media industry (Brand & Kinash, 2013). Even advanced courses on artificial intelligence (Bayliss, 2012) have been targeted. Mojang supports the strong interest of teachers through cooperation with the Finnish developer TeacherGaming LLC. They are using the modifiability of Minecraft to develop the teacher-friendly version MinecraftEdu (https://minecraftedu.com/). In this version of the game, easier creation tools for teachers to implement virtual assessments, simple server management, and tools to administer playing pupils (e.g., muting them, teleporting them or disabling inputs) are integrated. Thus, a vital community has developed (Minecraft Teachers, 2015), and teachers share their worlds (e.g., The Land of Turtles, where players learn to program small robots or the Wonderful World of Humanities where students explore different historical periods) with others around the world. Even educational videogames themselves can be a curricular topic by using Minecraft. Students learn what must be considered when creating educational videogames in seminars by developing short sections of gameplay with an educational focus and test these segments with pupils (e.g., students created an environment in Minecraft where children could recreate the Osterraziergang (Easter promenade) in Faust; Thiel, 2015).

How Minecraft is used in experimental research

The link between games and learning research has been investigated for several years. A small variety of games has been used to further differentiate this link. For example, the Squire’s Quest! was used to find out if healthy consuming behaviors among school children can be generated or changed by educational videogames (Baranowski et al., 2003). In a following project, these designers created another videogame, Escape from Diab (Thompson et al., 2008), in order to combine behavioral science and electronic games and prevent type 2 diabetes and obesity. Adams, Mayer, MacNamara, Koenig, and Wainess (2012) examined the effects of their educational games on learning processes within the domain of biology with Crystal Island and electro-mechanics with Cache 17. The Operational
Art Of War is an educational videogame designed by Frank (2012) that examines if the users’ focus on educational objective can be strengthened. With the help of the game Trade Ruler (Huang, 2011), a study examined how educational videogames might initiate and support learners’ goal-setting activities and influence their cognitive loads.

As long as researchers are technically experienced and have enough competence to construct their own videogames, it will be possible to investigate additional theories on learning and moderating variables within this area. However, this is not always the case. Most researchers experience that a simple modification within existing videogames is barely feasible as they cannot access the code of the game or not capable of modifying it. Therefore, it is very important to focus on modifiable games in research that allow even less computationally skilled researchers to expand their research focuses on educational videogames. This competency gap could be filled with Minecraft. Due to its numerous modifications, easy structure, huge player community, and countless forums, blogs, and YouTube videos on how to implement different features and rewrite source code, this game can be used even by game development novices.

Several studies that used this game as an excellent research “vehicle” can be cited. In the field of human-computer interaction and computer science, Orlikowski, Bongartz, Reddersen, Reuter, and Pfeiffer (2013) used Minecraft for research on jumping in virtual realities. Zorn, Wingrave, Charbonneau, and LaViola (2013) used the game to investigate how interest in programming can be increased by differentiating graphical code blocks from text code blocks. Marklund, Backlund, and Johannesson (2013) used different texture resource packs to track their players’ contributions (Figure 3) to discover collaboration tactics in emergent games in the field of game research.

![Figure 3. Comparison between color coded view and the player’s view (Marklund et al., 2013, p. 4)](image-url)

Working on the topic of instructional and educational psychology, Nebel, Schneider, and Rey (2016) used Minecraft in order to test if competitive groups’ conditions can increase the cognitive load as well as engagement, interest, and learning performances compared to the single player condition. The researchers built task sequences during parkour in the game world. Using the modification WorldEdit (https://github.com/sk89q/worldedit), the researchers easily duplicated the individual lanes to create a multiplayer scenario within a very short amount of time (Figure 4). Nebel, Schneider, Schledjewski and Rey (in print) used a huge world called Atlantis to examine if different goal settings theories can be confirmed within educational videogames. To establish a goal-free condition, a very open game
world was needed. In another experiment Nebel, Schneider, Beege, Kolda, Mackiewicz, and Rey (in print) utilized the free building mechanics to create an experiment addressing collaboration and task interdependence. The participants had to create a building of Fontane’s “Effi Briest” (2015) cooperatively with different sets of distributed building materials and layout information. As these examples highlight, Minecraft has proven to be a useful experimental research tool especially when open settings or large environments with many duplicate elements need to be addressed. Thus, it is capable of improving the technical repertoire of researchers and as a consequence supports research in the field of (educational) videogames.

Benefits of Minecraft

After a detailed analysis of the presented examples, several beneficial characteristics could be identified. Many of these benefits result from general gameplay mechanics. For example, blocks can be arranged in a way that could reproduce almost every static object or shape, thus providing stimuli for a very different set of education or research projects. Additionally, the unmodified game itself contains biomes that can be used as ecology representations (Ekaputra et al., 2013) and complex systems that can be influenced through the player. For example, Minecraft provides a simple ecological system with different forms of plants and animals. The player can use this system to create artificial crop farms and optimize this system for his or her benefit; thus, he or she participates in changing the environment. This is an excellent example of transformational play (Barab, Pettyjohn, Gresalfi, Volk, & Solomou, 2012; Barab, Gresalfi, & Ingram-Goble, 2010). The resulting accessibility enables active knowledge construction within constructivist approaches (Loyens & Gijbels, 2008; West & Bleiberg, 2013), as players can cooperate, self-regulate, and engage in problems with many interactive elements. Furthermore, the simple but ingenious idea of spatially represented blocks that provide simple functions (dirt can be worked, generate grass, can be further processed) can be extended with new functions or blocks (e.g., Codeblocks, Zorn et al., 2013). By using this mechanic, spatial, interactive knowledge creation is possible (Figure 5). Finally, the simple multiplayer structure of the game with individual servers enables collaboration between instructors (Short, 2012) and researchers in creating content or executing learning (or experimental) sessions, even within difficult worldwide projects (Mojang, 2015b) or special target groups (children with autism and their families, Duncan, 2015).

In contrast to these broad advantages, some benefits are closely connected to the actual version of Minecraft currently distributed and the technical dimensions. For example, transformational play can expand beyond the limitations of the original content, because modifications include more systems that could be influenced (e.g., quest mods, IndustrialCraft, MineChem). The borders of transformational processes are only limited by the Java-based framework as players can even crate modifications themselves. Despite the potential for learning, the modification friendliness offers further benefits. New gameplay elements can be included (e.g., zombies start to drop numbers upon their death, Al-Washmi et al., 2014) without interfering with the rest of the game and the need for comprehensive restructuring. Thus, there are low obstacles to include modifications as an learning topic (Bayliss, 2012). Even (simple) modification tools are available without any programming knowledge (i.e., MCreator, http://mcreator.pylo.si/). Researchers can use this simple content-creation mechanic to produce research environments or to gather data through plugins that track user information (Müller et al., 2014). Additionally, since a high number of players want to play (or learn) after the sessions (Förster, 2012) and Minecraft is a comparably cheap and technological simple application, many players can actually do so. This is especially important as teachers name licensing and expense as a major concern when they describe the barriers to using computer games at schools.

Figure 5. Different versions of visual coding (Left: spatial coding with code blocks. Right: visual turtle coding with ComputerCraftEdu)
(Williamson, 2009). Thus, the time-expensive strengths of Minecraft to enable freedom of the players within the game (Petrov, 2014) and to increase motivation to further explore the topic (Bayliss, 2012) can be given enough space. This advantage also applies to research projects that address long-term experiments, as there are comparably low obstacles for the test subjects to continue to play at home. Finally, the deliberately simple, stylized “rétro-chic” visual representation lowers the demands on the general appeal of Minecraft in comparison to the high visual standards of modern off-the-shelf videogames. Thus, the game could still appear aesthetically pleasing, even though modern games deliver far superior three-dimensional worlds. Additionally, the game can be updated and adjusted to the topic at hand through modifications, additional shaders and texture packs, and the simplistic visualizations can help focus the player on the intended topic.

Limitations of Minecraft

Despite the many benefits that led to numerous uses in education and research, Minecraft has several limitations within the topics teachers, players and gameplay/technical challenges. Focusing on the teacher side, although comparably simple, implementing Minecraft still requires specialists’ skills (Gregory et al., 2013) as necessary tools are not addressed in teacher qualifications. Therefore, the teachers’ engagement and skills have vastly different impacts on learning with the game. Some enjoy working with Minecraft, use their experience to react to unexpected situations, and extend this practice into other courses, while others need to work with clear guidelines (Hanghøj et al., 2014). Additionally, despite optimistic estimates (Ekaputra et al., 2013), a scenario is not automatically more fun by using Minecraft. Not every teacher or researcher is a talented game designer, and variations of the original Minecraft are not always entertaining to pupils (Petrov, 2014). Furthermore, including a multiplayer does not generate a social learning game per se since several harmful activities must be considered (e.g., hoarding resources, Hanghøj et al., 2014). Thus, researchers have to strictly monitor their experiments, in order to prevent data loss. Even if teachers or researchers put a lot of effort and thought into their projects, psychological questions need to be addressed. For example, teaching Boolean logic in a zombie-infested environment (Wingrave et al., 2012) might dampen learning outcomes.

There are also limitations regarding the players. Experienced players may fall back into their previous actions as the players may have experience with the game. Therefore, they might create and test less extensive new hypotheses (Dahlskog, 2012), or pre-existing ideas may persist even when the current learning environment contradicts them (De Jong, 2006). Skilled players may dominate in competitive or collaborative scenarios (Hanghøj et al., 2014; Marklund et al., 2013) or may be frustrated when the educational version of Minecraft lacks mechanics they are used to have. This impact of experience might be especially harmful when an experiment differs significantly from the original gameplay. In this case, it might be necessary to exclude very experienced participants or to include additional experimental time for players to become familiar with the new setting. Finally, the game’s original intent is based on free play with self-set goals using exploration and discovery. This aim increases the challenge of controlling players and setting functional borders — an aim that somehow violates the core principle of the game, especially for experienced players.

There are some downsides in gameplay and technical implementations as well. For example, the redstone mechanic in the game is user friendly (Ekaputra et al., 2013) but still requires an intensive learning phase before even simple mechanics can be created. Additionally, students have to deal with other challenges in addition to the specific learning task (e.g., leaning to move or to build in order to use redstone, or chunk size and implementation issues during AI course, Bayliss, 2012). In addition to the lack of a detailed physical system, the included block structure and the inaccurate fidelity of light are sometimes insufficient. For example, Minecraft lacks curved shapes and the specific light incidences for historical architecture, which are essential for recreating specific styles and archetypes. Other methods of three-dimensional content creation are currently in development (e.g., Landmark, Closed Beta, https://www.landmarkthegame.com/home) and show potential, although the delicate tradeoff between accessibility and creation potential must be considered carefully. Maybe an “advanced mode,” with simple but basic modifications such as rotation, might have a strong impact on the potential of Minecraft. Another challenge is the lack of a language feedback system (Hausrath, 2012) as detailed feedback is not in the intention of the game (e.g., the player freely decides what to do and what is “correct”). This increases the challenge of a successful implementation of tutorials on how to play the game or master the experiment. Therefore, teachers and researchers are often forced to use modifications to support learning with additional informative elements or interactive non-player-characters. This might lead to further technical difficulties in addition to the limited stability that results in slow gameplay or even
crashes if large numbers of blocks are affected (Bayliss, 2012). This is especially problematic in the case of experimental studies, because a loss of data or invalid test runs cannot be ruled out. Finally, otherwise desirable regular (client/Java) updates cause technical problems (Bayliss, 2012) since studies and courses often need long preparation times.

The future of Minecraft in education and research

Although a substantial interest in videogames in educational settings (Evans, 2014; Williamson, 2009) can be observed, further education of teachers is needed to ensure the proper use of Minecraft in educational settings and to enable the acquisition of cooperation partners in research. For example, students of the Technical University of Chemnitz tried to start a Minecraft project to teach literature. When they asked a teacher to participate in this project, she responded:

They do this [playing games] enough at home, and the parents are already struggling with that. […] Humans have and should work together as a team, but why should they take a diversion using a computer where everybody just sits in front of this box? […] I am a big supporter of the viewpoint that sometimes pupils have to learn and do something in school, no matter if it is fun or not. (translated from German)

This statement underpins the importance of spreading information about educational videogames, the use of Minecraft in general and the engagement in discussions with teachers and professionals in teacher education. Flexibility in lesson planning is important, as one session usually not enough time to master a topic completely. It is the responsibility of teachers and researchers to provide information about the social potential of educational videogames. With different viewpoints on the potential of videogames, very different reactions can be achieved. The response of a teacher presenting a Minecraft course emphasizes this argument:

We actually have some kids with mental health issues that just drive here! So, for example, a parent came in here with two sons, both with Asperger’s and they came in and used Minecraft together. And the parents, she stayed and watched and she was almost in tears because she said this was the first time that she saw an interaction with a teacher figure that was so positive that it was just letting them explore their interests and kind of learn as they go (Petrov, 2014).

Despite these social factors, the future of Minecraft in education and research is also influenced by its technical barriers and potential. Virtual worlds will increase in quality and fidelity (Gregory et al., 2013). Both factors are limited in Minecraft. However, this does not imply that Minecraft’s days are numbered. Instead, the lack of complex mechanics benefits approachability not only for students but also for teachers, one of the current driving forces behind the use of virtual worlds in education. Furthermore, the implication of Minecraft as a web-based tool (Walsh, Donahue, & Rhodes, 2015) could increase the distribution even more, as schools experience technical issues when they try to install necessary programs or run servers. This could be accompanied by deeper connection with mobile learning and the inclusion of Quick Response (QR) codes, hyperlinks, or Global Positioning System (GPS) coordinates. These supplements might benefit from the available mobile version of the game and could foster a connection with the physical world. New technical improvements within 3D printer projects (Haines, 2015) or connections with Raspberry PI, Arduino Leonardo, Lego We-Do, and Lego Mindstorms NXT will enable even stronger transitions to the real world. Additionally, the inversion of this connection is possible as well (e.g., interactive control of the virtual world through physical objects, Hanel, 2012). Furthermore, the transfer of geographic data to Minecraft has shown how real-world data can be processed and utilized for use in digital projects (e.g., Denmark recreated 1:1 scale in Minecraft, Høeg Nissen, 2014). This strong and technical comparably simple connection between virtual and physical spatiality could lead to further areas of applications, for example, perspective within the field of arts or social elements of geography.

Even without deep modifications or technical evolution, several effects, which have not yet been analyzed in educational videogames research, could be accessible by using Minecraft. For example, the seductive detail effect (Harp & Mayer, 1998; Rey, 2012) could be examined within videogames by easily modifying or placing some blocks in the game world or modifying NPCs with interesting dialogue. The theme and general appeal could be altered with the modification of texture packs, an approach that might be useful when addressing personalization or emotional design (Mayer, Fennell, Farmer, & Campbell, 2004; Moreno & Mayer, 2000; Schneider, Nebel, Pradel, & Rey, 2015). Another example might be the worked example effect (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Rourke & Sweller, 2009), since researchers can easily create different versions of the same learning content.
with different status of examples and solution progression. For example, a place to build logic circuits using redstone could be presented completely empty, with pre-build wires or with some gates already in the correct position. Increasing collaboration through the simple multiplayer mechanism (e.g., different parts of a redstone Von Neumann architecture collaboratively build within different educational facilities) might open up a new perspective for experiments on collaboration or embodiment research.

Implications from the success of Minecraft

Based on our experiences and the broad range of examples from education and research, we can finally derive implications from the immense impact of Minecraft in these fields. First, the simple exchange of worlds, creations, modifications, lessons, and experimental settings between teachers and researchers all over the world without strong software restrictions empowered the unintended use as an educational or experimental tool, tremendously. Users can not only share content but also modify the shared content further — as simple as writing a text document. This is fueled by simple content creation methods with representations of almost any static object. Especially for researchers in the field of videogames, Minecraft might be the simplest way to create custom experiments, where other games are usually lack modifiability. Another important lesson learned from Minecraft’s success is the game’s social impact on learners to become teachers themselves (e.g., Pithons, 2014). Fostered by the lack of guidelines and information and the freedom of creation, players engaged in long interactions with the game and its mechanics, creating hours of tutorial videos on YouTube and countless wikis. Thus, Minecraft paradigmatically shows a progression of learning that lasts longer than a one-time play: learning through playing (e.g., learning history with a game of hide-and-seek in virtual ancient Rome), learning through creating (e.g., expanding the city of Rome after searching for historical accurate information), and learning through teaching others (e.g., showing others the new version of Rome).

Due to the different functionalities of Minecraft and their applications, we propose interesting fields for the technical future of educational videogame software. Simple content creation using blocks or other simple shapes, combined with animation or physical functionalities, might be extremely interesting and result in powerful virtual environments. But the use of blocks as spatial functions, as successfully shown within Minecraft, might be even more important and could be expanded further. For example, with forms of simple, visual programming already used within tools like Kodu (http://research.microsoft.com/en-us/projects/kodu/) or Scratch (http://scratch.mit.edu/). To shed some more light on this point, we would like to present an example in Minecraft: ComputerCraftEdu (http://computercraftedu.com/) is a modification for Minecraft that includes turtles. These represent, simply put, a special block type that can be programmed like small virtual robots. ComputerCraftEdu uses a visual programming technique based on Lua (http://www.lua.org/) script language and thus enables the player to move the turtle around or build or mine blocks without advanced knowledge of informatics. The impact of a future educational and research tool would be amplified if this mechanic applied to all blocks. We can only begin to imagine what teachers, experimenters or players could create if they use visually programing within spatially distributable entities, group them, duplicate them, or create new types with functions reacting to various variables, like the player position, relation to other blocks or even events outside the game itself (e.g., GPS coordinates, external databases). As we have outlined, Minecraft offers tremendous qualities when it comes to creation, collaboration, and distribution and offers a wide range of pre-defined entities. Despite this, Minecraft lacks functionality when new entities are needed or when new functionalities have to be included by novices in programming. Therefore, overcoming these barriers requires a huge amount of creativity on the creators’ side and flexibility within the topics.

To finalize this review, and to indicate our suggested (technically idealistic) directions for further developments of creation tools in education and research, we want to briefly emphasize our three main implications: Content must be shared most easily, collaboratively created, and further modified without systematic restrictions or high technical demand. World building should be based on simple nucleus entities that can be created, modified, grouped, and organized. The spatially placed entities have to offer modifiable (ideally, with forms of visual programming) functions to enable creation or interactions with other entities, the player or other input from outside the game.

References


Nebel, S., Schneider, S., Beege, M., Kolda, F., Mackiewicz, V., & Rey, G. D. (in print). Increasing task interdependence in cooperative educational videogames to encourage collaboration.

Nebel, S., Schneider, S., Schledzewski, J., & Rey, G. D. (in print). Goal setting in educational video games: An Experimental comparison between goal-setting theory and the goal-free effect.


Owen. (2014, September 15). Yes, we’re being bought by Microsoft [Web log]. Retrieved from https://mojang.com/2014/09/yes-were-being-bought-by-microsoft/


Role-Play Game-Enhanced English for a Specific-Purpose Vocabulary-Acquisition Framework

Fang-Chen Lu and Ben Chang*

Graduate Institute of Learning and Instruction, National Central University, Taoyuan City, Taiwan // lufangchen7414@gmail.com // bchang.tw@gmail.com

*Corresponding author

(Submitted April 15, 2015; Revised July 9, 2015; Accepted September 3, 2015)

ABSTRACT

With the advantages of an engaged and authentic role-play game (RPG), this study aims to develop an RPG-enhanced English for specific purposes (ESP) vocabulary-acquisition framework, providing teachers and students a systematic way to incorporate RPG into ESP learning. The framework is composed of five parts: goal, three-level vocabulary sets, RPG contexts, a vocabulary test, and a participation questionnaire. In the perspective of three-level vocabulary sets, they are semantic sets, communicative sets, and situational sets. With regard to the vocabulary test, there are vocabulary assessment for semantic sets, vocabulary assessment for communicative sets, and vocabulary assessment for situational sets. In addition, the participation questionnaire classifies the students’ participation scale into personal interest, external expectations, social contact, and social stimulation. To evaluate the framework effectiveness, eighty northern Taiwan vocational high school second-year students, divided into an experimental group and a control group, were recruited to participate in this study. The results confirmed that the framework can facilitate students’ ESP vocabulary acquisition in vocabulary assessment for situational sets. Meanwhile, the framework was effective in promoting students’ social participation, especially on external expectations and social contact perspectives.

Keywords

RPG ESP framework, Semantic sets, Communicative sets, Situational sets, Students’ participation scale

Introduction

With the rapid development of business communities around the world, English-language skills have been considered an important ability needed to compete in the global economy. One of the goals of foreign-language education for vocational educational programs is to provide students with the foreign-language ability and advanced professional knowledge necessary to succeed in the job market. This development trend has caused English for specific purposes (ESP) instruction to be more greatly emphasized in the last few years at vocational high schools in Taiwan (Tsai & Davis, 2008).

ESP is well known as a learner-centered and content-based approach to teaching English as a foreign language that meets the needs of learners who need to learn English for use in their specific fields, such as business, science, technology, or academic learning (Hutchinson & Waters, 1987). However, there are some problems in the development of ESP courses. Porcaro (2001) indicated that authentic materials and specific knowledge were not provided in ESP courses. ESP has been looked upon as a neglected subject in vocational high schools because there is still room for improvement for ESP course design and implementation.

Authentic materials illustrate how English is used naturally by native speakers. Nunan and Miller (1995) defined authentic materials as those that were not created or edited expressly for language learners. One of the most challenging tasks constantly facing ESP teachers is how to bring authentic materials into the classroom. An advantage of role-play-game-based (RPG-based) learning, students are able to develop their own strategies for dealing with real language and stimulate their motivation to learn. Moreover, Tsai and Davis (2008) claimed that ESP concentrates more on language in context than on teaching grammar and language structures; therefore, authentic contexts and language use in typical work situations raise interest in learning ESP. RPGs, in which students are encouraged to come across different authentic materials related to their study subject, are highly graphical 2D or 3D video games played online, allowing individuals to interact not only with the gaming software but also with the avatars of other players through their self-created digital avatars (Yip & Kwan, 2006). The learning principle behind RPGs might be fact that thousands of participants can be online interacting with one another at the same time, engaging in a 3D online representation of actual locations (Childress & Braswell, 2006). Moreover, virtual learning
in an RPG environment provides a space for constructive learning and immerses learners in a meaningful communication simulated to authentic practices (Chuang, Chang & Chen, 2014). It is recommended that RPG be implemented in the contexts where students need to learn the knowledge and skills of English and practice them in authentic ways, as they provide active interactions and collaborations among learners and address cognitive issues and foster active learning.

### RPG_ESP Framework Goal: [Mission Statements & Need Description]

#### Using ChefVille* game to Facilitate ESP Vocabulary Acquisition

<table>
<thead>
<tr>
<th>Set / Icon</th>
<th>Feature</th>
<th>Voc. Assess.</th>
<th>Participation Scale Assess.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Semantic Sets</strong></td>
<td>S.: {word/synonym*, antonym}</td>
<td>Vocabulary assessment for semantic sets</td>
<td>Participation questionnaire for RPG_ESP vocabulary acquisition</td>
</tr>
<tr>
<td><img src="image" alt="Word/ Syn. Word*" /></td>
<td>Ex: [Word/synonym*, antonym*]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Word/ Ant. Word*" /></td>
<td>S1: {confirm/check*, deny*}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2: {purchase/acquire*, buy*}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3: {option/choice*, substitute*}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S4: {vegetarian/meat*}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S5: {ingredient/component*}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S6: {appetite/desire*, hunger*}</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Communicative Sets</strong></td>
<td>C.: [Communicative Session Description]</td>
<td>Vocabulary assessment for communicative sets</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="C1" /></td>
<td>{Sentence</td>
<td>S1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C1: {Sentence</td>
<td>S1, S2}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2: {Sentence</td>
<td>S3, S4}</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="C2" /></td>
<td>Ex: [Dealing with Complaints]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C1: I’d like to confirm my purchase order placed online yesterday.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2: But I could not find any option for a vegetarian.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3: Pardon me, but one of the ingredients is not for vegetarians and I’ve lost my appetite.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Situational Sets</strong></td>
<td>SS.: [Situational Session]</td>
<td>Vocabulary assessment for situational sets</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="SS1" /></td>
<td>a combination of Cj</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SS1: C1: confirm, purchase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C2: option, vegetarian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C3: ingredient, appetite</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="SS2" /></td>
<td>Ex: [A long and exhausting hike gave Mark a good appetite, so he ordered a big meal.]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M: Excuse me, madam.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W: Yes? What can I do for you, sir?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M: I’d like to confirm my purchase order placed online yesterday.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W: Do you have the reservation number?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M: Yes, here it is. But I could not find any option for a vegetarian.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W: (A while later.) My apologies, sir. Here is a new bowl of vegetable soup for you and a green salad, on the house.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M: Pardon me, but one of the ingredients is not for vegetarians and I’ve lost my appetite.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W: I’m awfully sorry, sir. I’ll have the chef cook you another one.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RPG Contexts: [RPG Environment Description]

Online Game: Restaurant settings of ChefVille

*ChefVille is an online RPG game available on Facebook for free*

Figure 1. RPG_ESP framework
The teaching of ESP vocabulary follows similar principles of English for general purposes (EGP); however, there is still room for ESP course teaching. The majority of educators consider the traditional lecture approach to ESP course implementation ineffective compared to active learning methods (Marbach-Ad, Seal & Sokolove, 2001). In addition, inappropriate instruction has the potential problem of leading to poor ESP course development (Kavaliauskiene, 2003). With the advantage of RPG-based learning environment, this study aims to develop an RPG-enhanced ESP vocabulary-acquisition framework, providing teachers and students a systematic way to incorporate RPG into ESP learning. Based on the framework, we further studied students’ ESP vocabulary effectiveness, and second, sought their participation scale for the framework in a Facebook online RPG game, ChefVille, to facilitate ESP vocabulary learning.

Role-play-game-supported English for specific purposes framework

To help students acquire more ESP vocabulary in an RPG learning environment, this study developed the RPG_ESP framework, which supplements students with ESP vocabulary acquisition. Figure 1 illustrates how RPG_ESP framework is applied to enhance students’ recognition and comprehension of ESP vocabulary. The framework aims to help students acquire more ESP vocabulary using organizing words to be learned in an RPG learning environment with the RPG_ESP framework availability based on semantic sets, communicative sets, and situational sets. The RPG_ESP framework may be advantageous to integrate the above three levels of ESP vocabulary acquisition in an RPG-enhanced learning environment. More specifically, this study proposed an RPG_ESP framework that contains five parts: the framework goal, three levels (semantic sets, communicative sets and situational sets to ESP vocabulary acquisition), RPG contexts, ESP vocabulary effectiveness, and students’ participation scale assessment. ESP vocabulary acquisition through three levels mentioned above increases their associative meaningfulness and is not easily forgotten. Nattinger (1980) argues that vocabularies need to be taught through techniques that organize vocabulary items to be learned. He stresses the importance of organizing items in terms of their meanings. Guirns and Redman (1986) point out that organizing lexical items semantically is very valuable in vocabulary learning. Similarly, Eyraud et al. (2000) also observe that knowledge of semantic relationship enables learners to exploit the meaning potential of words. That is, students can access the various shades of meanings of words, which would only be apparent when the words are studied in relation to each other.

The vocabulary test used in assessing students’ ESP vocabulary acquisition was comprised of three sections. Section I was a vocabulary assessment containing ten items covering words of semantic sets. Section II contained ten questions from communicative sets. In these two sections, there are twenty words randomly chosen from ChefVille’s daily questions. Section III included ten questions elicited from situational sets. ChefVille, as an instance of the framework, affords a platform to facilitate ESP vocabulary acquisition (as shown in gray in Figure 1). Thus, such a framework is expected to supplement students with recognition and comprehension of ESP vocabulary and motivate them to participate in an RPG-enhanced learning environment.

Semantic sets

Semantic sets contain words linked together by inferential relationships relating words together on the basis of meaning. Two common vocabulary techniques that link words in such a way are synonym and antonym. A synonym refers to words that are similar in meaning. Synonym as a technique has several advantages in vocabulary learning. It helps learners to learn the unfamiliar words in terms of the familiar. The meaning of an unknown word can be explained in terms of the word whose meaning is known. It also teaches learners the various lexical restrictions of the use of a given set of words. For instance, the words child and kid are synonyms, but they cannot be used interchangeably. The former seems to be more appropriate in formal situation, whereas the latter seems to be “right” in an informal situation. Similarly, we have positive and negative synonyms like thrifty and stingy; core and intensifier words such as mad and furious, which we may use depending on the kind of meaning we want to convey (Nattinger, 1980).

Antonym refers to words that are opposite in meaning. Antonym takes a variety of forms. The two basic forms are gradable and upgradeable opposites. Gradable opposites are opposites with degrees in between. An example of this kind of oppositeness could be the words hot and cold. The two words are gradable opposites because other words that can show different degrees of hotness or coldness can come between them, such as warm and tepid. Upgradeable
opposites, on the other hand, have a one-or-the-other kind of relationship. An example of upgradeable opposites is the oppositeness between the words male and female. The two words are not gradable because they are mutually exclusive (Atkins, Banteyirga & Mohammed, 1996). Antonym as technique plays a similar role in vocabulary learning to synonym.

**Communicative sets**

Communicative sets are words enclosed by semantic sets intended to deal with an authentic conversation situation (Canale & Swain, 1980). Another important function in communicative sets is for interpersonal relationship. Interpersonal skills are strategies employed by language users for facilitating language learning and enhancing communicative performance (Bialystok, 1990). As an individual interacts with his/her knowledgeable peers, learning becomes more supported, the user’s interest can be sustained, and s/he would get more involved in their learning and, thus, communicative competence can be evolved through the social interactions. Learning occurs when individuals are engaged in meaningful and authentic social activities (Vygotsky, 1978). The *zone of proximal development (ZPD)* is defined by Wertsch (1985) as “the distance between the child’s actual developmental levels as determined by independent problem solving and the higher level of potential development as determined through problem solving under adult guidance and in collaboration with more capable peers” (pp. 67–68).

**Situational sets**

Situational sets in vocabulary learning require the use of “general concepts” to bring together the specifics. They refer to the “cohesive chains” of relationships between words of communicative sets in a discourse (as shown in Figure 2). The technique stresses the need for using the context or situation in which words appear as an organizing principle. For example, cluster words such as “reservation,” “grilled salmon,” “salad dressing,” “appetite,” and “slow service” revolve around the central concept of “customer” in restaurant setting.

The organizing words to facilitate learning and recall is advocated by many researchers (Atkins, Banteyirga & Mohammed, 1996). Tinkham (1997) compared the learning of situational sets with that of sets of unrelated words in vocabulary learning. The results revealed that situation-related words—“library,” “whisper” and “quiet,” and “beach,” “sunny” and “swim”—were easier to learn than unrelated words—“fork,” “count” and “brave.” Thus, it is worthwhile for teachers to teach students vocabulary in specific situations.

Gairns and Redman (1986) also see another advantage of grouping conceptually related words together: it increases the chance for learners to draw on their knowledge of the world, making the process of learning and retrieval easier. Situational sets in vocabulary learning seem to be similar to what the model calls advance organizers, for it requires the use of “general concepts” to bring together the specifics. The example of a situational set was presented as the following (as shown in Figure 2).

![Figure 2. Situational sets](image-url)
RPG contexts: Restaurant settings of ChefVille

ChefVille is the game-based learning material that learners were assigned to use. ChefVille is one of Facebook’s online games; it lets users run their own restaurant, do the shopping, enquiry and purchase, make dishes, and serve the dishes to build up their unique restaurants (as shown in Figure 3). In addition, it may be a feasibly situated learning environment for problems solving and tasks completion in order to achieve a goal (as shown in Figure 4).

![Figure 3. Authenticity of the game ChefVille](image3)

![Figure 4. Situatedness of the game ChefVille](image4)

At first step, learners have to create their own character and then hire their friends as employees. The next step, they have to try best to collect required ingredients to level up different dishes in order to get more experience points. There are many ways to collect ingredients. For instance, they will get daily ingredients for logging in, passing the daily question, or completing the challenges and tasks. With these actions, the learners’ participation scale would be increased.

By accessing this vocabulary frequently for the necessary purpose of playing the game, learners could learn more vocabulary items related to a real-life situation. Additionally, learners can interact with others by visiting friends’ and others’ restaurants, leaving messages, and even trading ingredients with them to create a menu. Therefore, learners could learn more hospitality-related vocabulary by immersing themselves in the situation.

**Vocabulary-acquisition assessment**

The study estimated the size of participants’ current English vocabulary in order to compare it to their vocabulary acquisition when RPG_ESP framework was made available. The vocabulary test was administered as the post-test, with a total score of 100 developed for this study. The participants needed to take the vocabulary test comprising thirty items and was divided into three sections. Section I was a vocabulary assessment containing ten items covering words of semantic sets. Section II contained ten questions from communicative sets. In these two sections, there are twenty words randomly chosen from ChefVille’s daily questions. Section III included ten questions eliciting from
situational sets. The score obtained from the post-test was employed to identify their learning achievement. Figure 5 is an example for the testing items purposed to implement the effectiveness of ESP vocabulary acquisition using RPG_ESP framework.

<table>
<thead>
<tr>
<th>Part I: Vocabulary assessment for semantic sets</th>
<th>Part III: Vocabulary assessment for situational sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples:</td>
<td>Examples:</td>
</tr>
<tr>
<td><strong>1. Synonym:</strong></td>
<td>[A long and exhausting hike gave Mark a good appetite, so he ordered a big meal.]</td>
</tr>
<tr>
<td>(1) (pur)chase</td>
<td>M: Excuse me, Madam.</td>
</tr>
<tr>
<td>(2) option</td>
<td>W: Yes? What can I do for you, sir?</td>
</tr>
<tr>
<td>(3) ingredient</td>
<td>M: I’d like to <strong>(1) confirm</strong> my purchase order placed online yesterday.</td>
</tr>
<tr>
<td>(4) appetite</td>
<td>W: Do you have the reservation number?</td>
</tr>
<tr>
<td>(a) choice</td>
<td>M: Yes, here it is. But I could not find any option for a vegetarian.</td>
</tr>
<tr>
<td>(b) chef</td>
<td>W: (A while later.) My apologies, sir. Here is a new bowl of vegetable soup for you and a green salad, on the house.</td>
</tr>
<tr>
<td>(c) cashier</td>
<td>M: Pardon me, but one of the <strong>(2) ingredients</strong> is not for vegetarians and I’ve lost my appetite.</td>
</tr>
<tr>
<td>(d) desire</td>
<td>W: I’m awfully sorry, sir. I’ll have the chef cook you another one.</td>
</tr>
<tr>
<td>(e) component</td>
<td>M: Mark; W: Waitress</td>
</tr>
<tr>
<td>(f) buy</td>
<td>1. (a)check (b)book (c)contact</td>
</tr>
<tr>
<td></td>
<td>2. (a)components (b)inventories (c)requests</td>
</tr>
</tbody>
</table>

**Part II: Vocabulary assessment for communicative sets**

Examples:

1. ( ) How many ____________ are available for dessert on the menu?
   - (a)souvenirs (b)services (c)options

2. ( ) Exercise gave her a good ____________, so she ordered a big meal.
   - (a)appetite (b)apology (c)applause

... ...

**Figure 5.** Hospitality example for vocabulary test items

**Participation scale**

The questionnaire distributed to students revealed their participation scale for the framework in an RPG game ChefVille to facilitate ESP vocabulary learning. The questionnaire employed in the present study was adapted from the educational participation scale (EPS) proposed by Boshier (1991). A total of twenty items were retained in the final version of the survey and it has a proper reliability ($\alpha = .96$), suggesting that they had high reliability in assessing the high school students’ participation for Facebook online game ChefVille.

In addition, the questionnaire classifies the features of students’ participation into the following four scales: personal interest, external expectations, social contact and social stimulation. All items in the perceived participation degree were presented using a five-point Likert scale, from 1 (strongly disagree) to 5 (strongly agree). The details of the four scales are as follows:

- **Personal interest:** people who score highly on this scale participate in a Facebook online game, ChefVille, for their own interest. That is, they care about the inherent joy of the Facebook online game ChefVille, which impels their participation. A sample item of this scale is “I learn for the joy of it while participating in the Facebook online game ChefVille.”
- **External expectations:** people score highly on this scale participate in the Facebook online game ChefVille because of the expectations from someone at school. A sample item of this scale is “I participate in Facebook online game ChefVille because my peers encourage me to.”
• Social contact: people who score highly on this scale participate in the Facebook online game ChefVille because of the joy of interacting with others. A sample item of this scale is “I participate in the Facebook online game ChefVille to make more friends with the same interest.”

• Social stimulation: people who score highly on this scale are usually lonely or bored in regular life or teaching, and they participate in the Facebook online game ChefVille to meet others and to grapple with the problems in their social life. A sample item of this scale is “I participate in the Facebook online game ChefVille to take a break from my routine.”

Framework study

Research design

The study was designed to answer the two following research questions: (1) Do students actually acquire more ESP vocabulary using the RPG_ESP framework? and (2) What is the students’ participation scale for the framework in an RPG game, ChefVille, to facilitate ESP vocabulary learning? We investigated the first question by constructing and administering a vocabulary test reviewed by two high school English teachers to ensure the appropriate level for the participants. We investigated the second question via framing and implementing a participation questionnaire.

Procedure and instruments

The study purposed to provide a solution to the problem of limited ESP vocabulary acquisition if the RPG_ESP framework were made available. ChefVille is one of the English-language games on Facebook selected for display, and learners will learn vocabulary related to a real-life situation from this game. (https://www.facebook.com/ChefVille). The experimental process, shown in Figure 6, consisted of three steps. The instructions and grading criteria in the vocabulary test were stated to ensure that the participants truly understand what they were to be assessed on. The participants had thirty minutes to finish the vocabulary test, which consisted of thirty items, and was divided into three sections as a pre-test. A total of eighty participants were assigned to engage in the RPG environment approximately twenty hours spread over one month. The experimental group had access to the RPG_ESP framework, while the other forty participants did not. The game was regarded as participants’ additional English learning tool; therefore, they were expected to learn autonomously, with no stress, and acquire vocabulary spontaneously. After that, both groups answered the vocabulary test within thirty minutes as a post-test. Finally, the questionnaire was distributed to both groups.

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>Control group</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction &amp; pre-test administration</td>
<td>Introduction &amp; pre-test administration</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Logging in to the Facebook online game ChefVille with RPG_ESP framework availability</td>
<td>Logging in to the Facebook online game ChefVille without RPG_ESP framework availability</td>
<td>20 hours</td>
</tr>
<tr>
<td>Post-test administration</td>
<td>Post-test administration</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Questionnaire distribution</td>
<td>Questionnaire distribution</td>
<td>20 minutes</td>
</tr>
</tbody>
</table>

*Figure 6. Experimental flow*

Data collection and analysis

To answer the first research question, an independent-samples t-test was conducted to answer the question, “Do students acquire more ESP vocabulary using the RPG_ESP framework?” The purpose for this analysis was to
compare the performance between the experimental group, who were assigned to engage in the RPG environment with RPG_ESP framework availability and the control group, without RPG_ESP framework availability. The vocabulary test was used to demonstrate the potential quantitative differences between ESP vocabulary acquisition facilitated with and without the RPG_ESP framework availability. The participants had thirty minutes to finish the vocabulary test of thirty items divided into three sections. Section I was a vocabulary assessment containing ten items covering words of semantic sets. Section II contained ten questions from communicative sets. In these two sections, there are twenty words randomly chosen from ChefVille’s daily questions. Section III included ten questions eliciting from situational sets.

Next, information collected from the questionnaire was used to answer the second research question, “What is students’ participation scale for the framework in an RPG game, ChefVille, to facilitate ESP vocabulary learning?” The questionnaire comprising twenty items classified the features of students’ participation into the following four scales: personal interest, external expectations, social contact and social stimulation. In addition, the proportions of the selected options of each question were calculated and used as the basis for analyzing and reporting the students’ participation scale for the Facebook online game ChefVille in assisting ESP vocabulary acquisition.

Results and discussion

Situational sets effect

The participants of this study were eighty students from two classes of an urban vocational high school in Taiwan. Table 1 provides correlations between the two paired scores of the pre-test and post-test. The correlation ($r = .77$) between the pre-test and post-test indicates that learners who scored high on the pre-test were very likely to score high on the post-test, and learners who scored low were very likely to score poorly on the post-test. More specifically, it means that the vocabulary test has reliability, which systematically measures primarily the same thing both times it is taken.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RPG_ESP framework</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary Test</td>
<td>With</td>
<td>40</td>
<td>78.33</td>
<td>12.47</td>
<td>.01*</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>70.23</td>
<td>14.17</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RPG_ESP framework</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary assessment for semantic sets (up to 30%)</td>
<td>With</td>
<td>40</td>
<td>21.03</td>
<td>7.33</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>20.40</td>
<td>7.60</td>
<td></td>
</tr>
<tr>
<td>Vocabulary assessment for communicative sets (up to 30%)</td>
<td>With</td>
<td>40</td>
<td>23.40</td>
<td>6.70</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>22.13</td>
<td>7.05</td>
<td></td>
</tr>
<tr>
<td>Vocabulary assessment for situational sets (up to 40%)</td>
<td>With</td>
<td>40</td>
<td>33.90</td>
<td>6.34</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>27.70</td>
<td>10.22</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05

Before this study, these two groups did not show statistical difference in their English academic achievement ($p > .05$). As shown in Table 2, the results indicated that the participants in the group of ESP vocabulary acquisition with the RPG_ESP framework outperformed those ESP vocabulary without the RPG_ESP framework in ESP vocabulary test ($p < .05$). Noteworthy, as shown in Table 3, statistically significant difference was found in Part III: Vocabulary assessment for situational sets but not for the other two parts, namely Part I: Vocabulary assessment for semantic sets and Part II: Vocabulary assessment for communicative sets across the two learning conditions between
the experimental group: ESP vocabulary acquisition with the RPG_ESP framework, and the control group: ESP vocabulary without the RPG_ESP framework ($p < .05$).

**Participation scale effect**

To validate the questionnaire and clarify its structure, an exploratory factor analysis with a varimax rotation was performed. Items with a factor loading of less than 0.50 and with many cross-loadings were excluded from the instrument (Costello & Osborne, 2005). Through the factor analysis, the final version of the survey consisted of twenty items, which were retained in four scales. Furthermore, the reliability coefficients for the scales respectively were 0.80 (personal interest, five items), 0.80 (external expectations, five items), 0.90 (social contact, five items) and 0.90 (social stimulation, five items), respectively, and the overall alpha was 0.96, suggesting that they had high reliability in assessing the high school students’ participation in the Facebook online game ChefVille.

Table 4 shows students’ means and standard deviations of responses to each scale assessed by the survey developed in this study. As shown in Table 4, the students scored highest for both two groups on the social contact scale ($M = 4.58$ and $3.95$ on a five-point scale, respectively), followed by the social stimulation scale ($M = 3.70$ and $M = 3.90$, respectively), the personal interest scale ($M = 2.73$ and $M = 2.93$, respectively), and the external expectations scale ($M = 2.98$ and $M = 2.03$, respectively). The students scored highest on the scales of social contact and social stimulation, suggesting that the students’ social interaction experience may facilitate their participation scale for the RPG game ChefVille.

Table 4. Participation scale for RPG_ESP vocabulary-acquisition survey

<table>
<thead>
<tr>
<th>Factor</th>
<th>RPG_ESP Framework</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1: Personal interest</td>
<td>With</td>
<td>40</td>
<td>2.73</td>
<td>1.38</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>2.93</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>Factor 2: External expectations</td>
<td>With</td>
<td>40</td>
<td>2.98</td>
<td>1.31</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>2.03</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Factor 3: Social contact</td>
<td>With</td>
<td>40</td>
<td>4.58</td>
<td>0.64</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>3.95</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>Factor 4: Social stimulation</td>
<td>With</td>
<td>40</td>
<td>3.70</td>
<td>1.31</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>Without</td>
<td>40</td>
<td>3.90</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>

Note. *$p < .05$.

According to Bandura (1986), students with successful learning experiences are the most influential source in the development of learners’ beliefs. Students are likely to imitate the behavior of those they believe are competent. Also, the findings derived from recent research have suggested that exposure to competent students with successful experiences ensures success in building positive students’ beliefs about learning (Siwatu, 2007). The framework seems to suggest a possible way for promoting students’ willingness to share their experiences regarding ESP vocabulary acquisition as well as motivate them to participate in an RPG-enhanced learning environment.

**Conclusion and future research**

This study investigated the effectiveness of ESP vocabulary acquisition using the RPG_ESP framework and the students’ participation scale for the framework in an RPG game, ChefVille, to facilitate ESP vocabulary learning. In accordance with the results and discussion, several major conclusions are made. First of all, the results indicated that the RPG_ESP framework was beneficial to students’ ESP vocabulary acquisition only with respect to the contextualized use of words. A possible explanation for the result was that RPG_ESP framework could efficiently situate learning in an authentic context and lead learners to learn via the meaningful cognitive process. Lewis (1993)
asserted that the authentic context makes language useful at once to learners because they can apply diverse and pragmatic language in daily life. Additionally, the reason why they had better ESP vocabulary acquisition was because the RPG_ESP framework presents vocabulary in a systematic way that is beneficial for learning. In terms of the participation scale, the students scored highest on the scale of social contact followed by social stimulation, indicating that the sociocultural approaches linking to students’ daily life experiences as well as improving their social interaction with peers might be able to foster the students’ participation scale for the framework in the RPG game ChefVille to facilitate ESP vocabulary learning. In addition, the two scales of personal interest and external expectations have to do with students’ past successful experiences or positive authentic accomplishments, which could increase their participation scale in an RPG game.

The findings of this study offered three important implications for future research. First, the results indicated that the RPG_ESP framework facilitated students’ ESP vocabulary acquisition. This finding may provide implication for teachers or educators when the RPG_ESP framework availability was added for EFL students in terms of self-efficacy and learner autonomy. This is done with the hope that it may provide an alternative solution to the problem of limited ESP vocabulary learning and teaching. This study would enable educators and learners to realize and consent to the view that the RPG_ESP framework could contribute to ESP vocabulary acquisition.

Additionally, it was found that RPG_ESP vocabulary acquisition linking to students’ daily life experiences might be the key component of their participation. Alemi (2010) proposed the role of using games in expanding the learner’s vocabulary as well as motivation. It is important because we can better know about learners’ perspectives to achieve learners’ actual needs. Also, based on learners’ points of view, it can assist educators in adopting a suitable education model to fit individual differences and refine pedagogical approaches as well.

Finally, this study proposed an RPG_ESP framework containing three levels: as semantic sets, communicative sets and situational sets. Such a framework was expected to supplement students with recognition and comprehension of ESP vocabulary and motivated them to participate in RPG-enhanced learning environment. In order to construct authentic contexts and language used in the workplace, the RPG_ESP framework could help ESP teachers design teaching materials and evaluation tasks to familiarize students with the real-life workplace, which exactly meets learners’ needs. In sum, the findings of this study provided valuable empirical evidence for classroom teachers to enhance the effectiveness of ESP vocabulary teaching and learning.

Some of the limitations of this study could serve as future directions for conducting related studies. First of all, the newly developed participation scale for RPG_ESP Vocabulary-Acquisition Survey should be validated with a larger sample across different grade levels and with a more rigorous method such as confirmatory factor analysis. Moreover, as previously mentioned, researchers can adopt qualitative or mixed methodologies to explore students’ participation scale for the framework in the RPG game ChefVille to facilitate ESP vocabulary learning from different perspectives and identify additional scales. Third, the structure type of the vocabulary test was all in a multiple-choice format. Wittgenstein (1969) claimed that the meaning of a word is its use, and words are how you use them. Wittgenstein’s view of language as social practice is instructive for anyone who seeks to communicate clearly and effectively. However, the test scope was narrowed down to gain more hospitality-related vocabulary only, which might be problematic in revealing students’ pragmatic language production in real life.

References


Enacting Viewing Skills with Apps to Promote Collaborative Mathematics Learning

Kay Yong, Khoo
E-Learn Dot Com Limited, Yuen Long, NT, Hong Kong // kykhs@yahoo.com

(Submitted April 21, 2015; Revised September 9, 2015; Accepted September 15, 2015)

ABSTRACT
This paper comprises discussion on the research findings of this study into how apps can be used in the classroom to collaboratively promote construction of mathematical knowledge in children in ways that fundamentally transform the instructional environment. The study results identify how children enact viewing skills through digital texts to acquire new knowledge in their addition and subtraction learning. These skills enable children to externalise their understanding and internalise new meaning-making when interacting with peers. The problems presented to the participants were non-routine and immersive through externalising understanding with behavioural intensity and emotional quality at optimal levels. The subsequent enactment of the viewing skills led to the internalisation of the new knowledge. However, these dual reciprocal learning approaches require due consideration of elements of the learners’ personalities, the rules of the learning games, and the community settings of the classroom, all of which are crucial in determining the learners’ engagement in a learning activity and active involvement in associated learning processes. A detailed examination of the meaning-making processes through which viewing skills mediate children’s knowledge acquisition while seamlessly switching between individual and social interactions has led to the development of the framework of the preschool classroom self-directed mathematics learning.

Keywords
Viewing skills, Problem solving, Creative reasoning, Collaborative learning, Mathematics

Introduction
Mathematics competencies are cumulative over time (Jordan, Kaplan, Remineni, & Locuniak, 2009; National Mathematics Advisory Panel, 2008). If not properly addressed and overcome, difficulties encountered at any stage of learning will lead to poor achievement in subsequent mathematics learning. Problem-solving skills in addition and subtraction are dominant aspects of the fundamental competency domains in mathematics. When children practise solving problems, their underlying conceptual and procedural knowledge in addition and subtraction determines their competency (Canobi, Reeve, & Pattison, 2003; Canobi, Reeves, & Pattison, 1998). The causal relations between these two areas of knowledge have been found to be bi-directional; increases in conceptual knowledge will help to increase procedural knowledge, and vice versa (Rittle-Johnson & Schneider, 2015). Therefore, in developing an effective classroom teaching pedagogy, the integration of content knowledge (conceptual and procedural) into learning approaches is important. These approaches must engage children in the target-learning activities while leading to the process of meaning-making using the content knowledge (Hiebert & Wearne, 1996; Schneider, Rittle-Johnson, & Star, 2011).

The conceptual and procedural knowledge of learners is observed explicitly via strategies applied during their routes to problem solutions. Some of the strategies applied in addition problem solving are direct modelling (represented by objects, which are all counted), counting on from the first number, counting on from the larger number, and recall with no apparent counting; those applied in subtraction problem solving are direct modelling (counting objects by separating from the total and counting those remaining), counting down from the bigger number (a backward counting sequence from bigger numbers), and counting up from the smaller number (a forward counting sequence from smaller numbers) (Carpenter & Moser, 1984).

Learning by rote involves imitating. Conversely, creative reasoning engages students by allowing them to develop well-founded mathematically anchored arguments for their choice of methods in non-routine problem-solving processes. In conjunction with challenging non-routine problems, collaboration is often suggested because it can improve students’ conceptual understanding (Boaler & Greene, 2000; Stahl, Rosé, & Goggins, 2011). However, to accomplish collaborative creative reasoning, a suitable learning environment needs to be established. In this learning environment, students need to apply new strategies repeatedly, with the objective being the advancement of their
competency in addition and subtraction problem solving. Collaboration on a challenging problem cannot be automatically initiated within groups. The utilisation of the process of negotiation to seek the new knowledge (i.e., the correct strategies) must be made visible to the learners during their collaborative efforts. Therefore, mathematics learning that is solely print-based and structured by content printed in a book is inadequate (Clausen-May, 2013). In terms of teaching and learning technology resources, there are a number of readily available free online mathematical problem-solving digital apps. These digital apps mainly afford opportunities to learn interactively with ideas, content and modalities that were not previously possible (Yelland, 2015). However, the place of these apps in formal education in kindergarten has been contentious (Zaranis, Brown, Evans, & Hannula, 2013).

It is believed that working collaboratively with these digital texts may allow participants to jointly acquire knowledge none of them had before (Cress, Stahl, Ludvigsen, & Law, 2015). However, the use of well-established apps and environments alone does not guarantee that collective knowledge construction will take place. That said, teaching and learning with digital apps could create and facilitate learning contexts but not the actual learning. Previous studies and review of the theoretical constructs have highlighted the importance of determining how digital texts facilitate learning when the apps are collaboratively used.

Interaction with digital text incorporates the viewing skill of the participants, which enables them to select an element (of different modes or with a unique digital functionality) and read the information pertaining to the element, then proceed to read the next element. In the course of engaging with the elements, they will integrate them (spatially or temporally) to make meaning (Khoo & Churchill, 2013) (see Table 1). Learning mathematics and acquiring the competency to solve problems have been largely understood as rational cognitive processes (Zan, Brown, Evans, & Hannula, 2006). The actual learning takes place when learners make sense of mathematics through a meaning-making process. The process of meaning-making is implicit and indirect (Seeger, 2011). There is an emerging gap between the capabilities of digital texts in meaning-making and how the participants appropriate digital apps in their collaborative mathematics learning. Therefore, the focus of this study is not students’ learning, but rather on how, during their collaboration, viewing skills facilitate instant immersion in mathematical problem-solving practices.

**Table 1. The framework of viewing and representing skills (from Khoo & Churchill, 2013)**

<table>
<thead>
<tr>
<th>Macro process</th>
<th>Element selection</th>
<th>Element integration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skills</strong></td>
<td><strong>Multi-mode</strong></td>
<td><strong>Contextual link</strong></td>
</tr>
<tr>
<td></td>
<td>The skill to recognise and create elements of different modes to form information.</td>
<td>The skill to interpret and create contextual links (in spatial/temporal layouts) with different elements to form information.</td>
</tr>
<tr>
<td></td>
<td><strong>Affordances of modes</strong></td>
<td><strong>Navigation</strong></td>
</tr>
<tr>
<td></td>
<td>The skill to apply and engage with the affordances of different modes in elements to form information.</td>
<td>The skill to move around a screen to integrate different elements to form information.</td>
</tr>
<tr>
<td></td>
<td><strong>Digital functionality</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The skill to assimilate digital functionalities in elements to form information.</td>
<td>The skill to assimilate digital functionalities to integrate elements to form information.</td>
</tr>
</tbody>
</table>

**Research design**

The current study seeks to analyse how children acquire new knowledge through activities undertaken in the context of collaboratively using apps. The activity cannot be understood outside the context in which it occurs. Analysis of human activity must take into account the participants and their goals, the outcomes resulting from their activity, and the rules that circumscribe the activity and the community in which it occurs. The activity theory (Engestrom, 1987) was selected as the theoretical framework best applicable in the contexts of learning studied here (see Figure 1).

The research design is qualitative. To gain a more robust and generalisable result, we adopt a multiple case study research approach to independently confirm emerging constructs and propositions, while at the same time revealing complementary aspects of the phenomena of interest (Yin, 2014). The study included observations, video recordings and interviews, and employed an inductive research strategy that intuitively developed abstractions from the research
Two research questions emerging from the literature reviews guided the data collection and analysis of the current study:

1. How do digital texts facilitate collaboration in mathematics problem-solving activities?
2. What elements of digital-based problem-solving activities might contribute to or obstruct students’ collaborative creative reasoning practices?

In order to gain insight into the phenomena of interest as incorporated in the research questions, we need to apply a series of research approaches. The phenomena were derived from natural settings that are conflated with both the unit of study and the product of investigation. Data was collected over a period of eight months. The study was explorative, and a constant comparative method of data analysis was employed (see Table 2).

---

**Figure 1.** The research framework of the study

---

**Table 2. Research methodology procedure and aims**

<table>
<thead>
<tr>
<th>Data Collections</th>
<th>Observation</th>
<th>Classroom observations</th>
<th>Interviews with the teachers</th>
<th>Analysis of the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collections</td>
<td>To find out whether:</td>
<td>-To observe how the participants learn in the classroom with mobile devices.</td>
<td><strong>Unstructured interviews</strong> -To determine the objectives of using the mobile devices in the classroom.</td>
<td>To study how the participants enacted digital texts with apps to gain new knowledge.</td>
</tr>
<tr>
<td></td>
<td>-The kindergartens were teaching mathematics with mobile devices.</td>
<td>-To study how the participants interacted with the artefacts, their peers, the community and the rules in the lessons.</td>
<td>-To identify the addition and subtraction strategies taught by the teacher to the participants.</td>
<td>-To study how the participants interacted with their peers in lessons and understood the rules, and the math concepts in the artefacts.</td>
</tr>
<tr>
<td></td>
<td>-The children used the devices individually and received appropriate feedback in the process of learning, resulting in positive learning taking place (i.e., they acquired new knowledge).</td>
<td>-To video-record the lessons.</td>
<td>-To study how the teachers mediated mobile technology in learning.</td>
<td>-To investigate how the emerging skills mediate digital text in learning addition/subtraction.</td>
</tr>
</tbody>
</table>

**Key Questions**

- Kindergarten(s) that applied mobile learning in mathematics resulting in authentic learning being observed were shortlisted.
- To determine how the students used viewing and representing skills emerging from use of mobile devices.
- To ascertain the social learning skills emerging from use of mobile devices.
- To verify the researcher’s observations on how the teachers mediated mobile technology in learning.
- To confirm the observational data during the lessons.

| 1st – 4th months | 5th – 8th months |
The study commenced by selecting suitable participants (the sampling units). The selection was terminated when no new information was forthcoming from sampled units during the research period. Four participants were identified according to their personalities and their community settings in the classrooms: their profiles are presented in Table 3. Pseudonyms are used to preserve their anonymity.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Gender</th>
<th>The lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter</td>
<td>5</td>
<td>Male</td>
<td>Team A</td>
</tr>
<tr>
<td>Mary</td>
<td>5</td>
<td>Female</td>
<td>Each participant was provided with one device. They practiced individually.</td>
</tr>
<tr>
<td>Ben</td>
<td>5</td>
<td>Male</td>
<td>Team B</td>
</tr>
<tr>
<td>Nicole</td>
<td>5</td>
<td>Female</td>
<td>Each participant was provided with one device. They practiced individually.</td>
</tr>
</tbody>
</table>

Each participant performed in front of the group until they had completed their practice.

### Table 3. Details derived from the selection criteria for the four cases

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Gender</th>
<th>The lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter</td>
<td>5</td>
<td>Male</td>
<td>Team A</td>
</tr>
<tr>
<td>Mary</td>
<td>5</td>
<td>Female</td>
<td>Each participant was provided with one device. They practiced individually.</td>
</tr>
<tr>
<td>Ben</td>
<td>5</td>
<td>Male</td>
<td>Team B</td>
</tr>
<tr>
<td>Nicole</td>
<td>5</td>
<td>Female</td>
<td>Each participant was provided with one device. They practiced individually.</td>
</tr>
</tbody>
</table>

Each participant performed in front of the group until they had completed their practice.

### Table 4. The learning apps

#### Addition & Subtraction for Kids (the first app)

There are ten questions in a set. The numbers featuring in the questions were from 1 to 20 and were randomly set. Each question required the use of either addition or subtraction. No time limit was set for the questions. The accumulated scores and number of wrong attempts appeared at the top of the screen.

#### Math Adventure (the second app)

There were five different themes (Penguin Craze, Winter Match, Catch a Star, Snowman Hunter and Math Bingo). Different themes provide different instructions in problem solving. The questions had to be completed within a time limit.

A. Penguin Craze (addition) allows the participants two lives (two chances to make mistakes before the game is terminated) and lasts for a maximum of three minutes.

B. Catch a Star (subtraction) lasts for a maximum of three minutes.

C. Winter Match (addition) has 20 boxes, 10 containing numbers, 10 an incomplete addition equation. The player clicks a box and subsequently selects an answer by clicking another box. The score reduces as time advances, reaching a zero score in 110 seconds.

#### Juicy Math (the third app)

The game provides three choices: addition only, subtraction only or both. There is no set time limit for completion of the questions. The screen displays three questions at any one time.

#### Worm Jump (the fourth app)

The bird in the game moves forward a step at a time. The worm is ahead of the bird and once a question is answered correctly, the worm moves forward a step. If the answer is incorrect, the worm will stay put. If the bird reaches the worm, the worm will be eaten.
During the eight months’ study at the kindergarten, the teacher used different mathematics apps in children’s practices in the classroom. The apps were introduced after the formal mathematics lessons. Each of the apps was used repeatedly two to four times per week. All the children were given the opportunity to use the apps during practices. Four apps were selected for the current study: Addition & Subtraction for Kids, Math Adventure, Juicy Math and Worm Jump. These apps were selected based on the unique characteristics of the instructions they applied in learning mathematics (see Table 4).

**Observations of participants enacting viewing skills when using app**

**Participant one: Peter**

Peter was an outgoing child, asking questions spontaneously of his classmates in the classroom whenever doubts arose in his mind. His recent math assessment result was about average. In Peter’s engagement with the first app, two students were assigned one mobile device (see Table 3) and took turns to answer the questions. The teacher briefed all the children in the class before they started with the strategies of “counting on from the first number” and “counting down from the last number” for solving both addition and subtraction questions. Peter started the game before Mary. In interacting with the first question (6 – 1 = ? [see Element 1 in Figure 2.1]), the researcher observed that he counted all the fish and omitted one from the total with his finger—that is, using direct modelling (see Element 2). He then selected the answer, “5” (see Element 3). The app prompted the “right” response, which he acknowledged with a smile. After attempting the question, Mary explained the strategies as briefed by the teacher. Subsequently, Peter answered the remaining questions by applying the new strategies. He turned to Mary for confirmation each time he had made a tentative choice of answers. Mary nodded to indicate her agreement with his choices. He was slow and doubtful in his first attempt to apply the new strategy but subsequently demonstrated competency in its application. The situation demonstrated the process of peer learning (Liu & Carless, 2006) mediated in the context of enacting with digital texts.

![Figure 2. Peter enacted the viewing and representing skills](image)

The following snippets of the interview with Mary were conducted after the lesson.

The researcher: “Mary, why did you tell Peter to solve the question by counting down from 6 by one?"

Mary: “This is the method that the teacher told us.”

The interview and the observation suggested that Peter enacted viewing skill (Khoo & Churchill, 2013) by selecting one element and understood the task at hand, i.e., 6 – 1. Subsequently, he selected another element (the fish) and counted them. He proceeded to select answer “5” at the bottom of the screen. The digital function of the app responded with “RIGHT.” He integrated the information of the three elements to make meaning. The enactment was observed and understood by Mary.
For the second app, Peter was assigned a personalised device. He selected the Winter Match in his attempt to solve the set problems. A countdown timer limited to 110 seconds the time Peter was given to finish the questions. After this time period had elapsed (see Element 2 in Figure 2.2), the game could still be continued but the score was always zero (see Table 4). The teacher set a rule that each of the children had to finish all the questions before the score turned to zero, as per Liu and Carless’ (2006) assertion that “engaging learners in thinking about achieving outcomes to certain agreed criteria is a learning process” (pp. 280). The researcher observed that Peter was aware of the criteria of the game. He started by selecting the first box (see Element 1) that is, “5 + 5,” upon which the colour would change to blue. He would then pause and use his fingers to count (1, 2, 3, 4, 5 on one hand and 6, 7, 8, 9, 10 on the other), following which he concluded that the answer was “10.” He was applying the strategy of modelling. He would then select the box “10.” The app would respond with a “well done!” notification, and the two boxes would close. He continued until the sixth question but the score dropped to 0 before he was able to complete all ten questions. Peter was enacting viewing skills by selecting elements of different information and integrating them to form meaning. He showed his excitement each time he completed a question. This was in keeping with Dawley and Dede’s observation (Dawley & Dede, 2014) that an immersive sensation is achieved through the use of sensory input with digital texts and digital functionality.

Peter sought feedback from the teacher each time he was in doubt. The teacher guided him from time to time by introducing the strategy of “counting on from the first number.” Slowly, after a few practices, he switched to this strategy. He achieved the score of 100 on his tenth attempt, having correctly completed the ten questions. He demonstrated his artfulness in engaging with the learning context, while his impromptu interactions with the teacher and the teacher’s feedback regulated his learning. His ability to seek feedback was observed. He managed to learn a new strategy during his practices with the second app. The interview after the lesson confirmed that he was aware that it was critical to change the strategy to achieve the score of 100.

In engaging with the third app, he was also given a personalised mobile device; the instructions of the app provided neither scores nor time limit. The researcher observed that Peter attempted the questions using the strategy of “counting all” with the objects in the two boxes. He applied the same strategy to the rest of the questions. He was unable to obtain the answers for some of the questions in his initial attempts.

Peter played the game featuring in the fourth app in front of the teacher and his group of classmates. It was a 40-minute lesson. The teacher set the condition that each child should answer at least ten questions before the termination of the game and would have to repeat it if fewer than three answers were incorrect. The researcher noted that Peter started his first attempt with the “counting on from the first number” strategy. He completed the next two questions but the game was terminated before he could finish the fourth. He showed his disappointment when the bird ate the worm. Peter observed how other children completed their attempts. The group was noisy. Some would speak out the numbers and answers without any apparent attempt at counting: e.g., 3, 5 [pause]... 8. When it was Peter’s turn to attempt the questions, it was observed that he spoke the questions aloud and followed by answering aloud spontaneously without counting on his fingers. His short response time enabled him to succeed on his third attempt with a score of 120 (twelve correct answers), thus completing ten questions consecutively. He also acquired the number fact strategy by practising the viewing skills with digital texts and retrieving the recalled number fact from his memory (Carpenter & Moser, 1984). The metaphoric animations created situations that enabled him to learn a new strategy, and a learning hierarchy from acquisition to fluency was observed. Further, his observational skill in learning was evidenced (Browder, Schoen, & Lentz, 1986).

Participant two: Mary

Mary had an easy-going manner, and demonstrated good social assertive skills in the class. She had scored good results in her schoolwork in mathematics and was very helpful to her classmates when asked for assistance. As with Peter, Mary demonstrated competencies in engaging with digital texts on screens and in meaning-making by applying appropriate viewing skills.

Prior to her engagement with the first app, she had had a vicarious experience from teaming with Peter and guiding him to complete the questions. She completed the first and the third apps promptly. While she demonstrated her ability in enacting viewing skills in both the apps, there were no signs that she had learned a new mathematical strategy. However, in her second app attempt, she was provided with an individualised mobile device to operate on
her own, and given the same rules as Peter on how to complete the task. However, the researcher observed that she started with a different approach: she clicked two consecutive buttons with the same values to close the two buttons (i.e., she clicked 4 & 4; 5 & 5 etc. instead of 1 + 3 then 4; 2 + 2 then 4; 1 + 4 then 5; 2 + 3 then 5). She managed very quickly to complete five questions. Each time, the app reacted with an encouraging “Well done!” to which she responded with a smile. When she attempted the sixth set of questions, which were of an unfamiliar problem-solving nature, she was panicked by the remaining buttons (these being 4 + 2, 1 + 5, 2 + 2, 1 + 3, etc.). She paused and pondered, then sought assistance from her teacher, who duly clued her up. The teacher introduced a new strategy, the “recall number fact” with no apparent counting (Carpenter & Moser, 1984) and repeated the same strategy with the two subsequent questions. The interview with Mary confirmed that she had wanted to complete the questions as soon as she could.

The researcher: “Why did you select 4 and 4, 5 and 5, etc., initially?”
Mary: “It was a fast way to complete the questions.”

As Mary enacted viewing skill to solve the questions, she created a problem for herself when she reached the sixth question. We observed how the teacher gave the hints to help her to apply the new strategy to complete the task. The scaffolding process was evidenced (Van De Pol, Volman, & Beishuizen, 2010). In the fourth app, with the assistance of answer hints from the teacher and her peers, she demonstrated her ability to answer the two-digit and one-digit questions with a new strategy, the decomposition strategy (e.g., $\frac{13 + 5}{10 + 3 + 5} = 18$) (Canobi et al., 1998), learning through reflection from the teacher and peer hints. In her attempt with the fourth app, she was observed to demonstrate her ability to learn the new strategy of recalling number facts from memory through her observation of the group activity.

**Participant three: Ben**

Ben was a brilliant child. He had achieved 100% scores in most of his schoolwork. He was observant, quiet and able to learn quickly with less practice and repetition than his peers. When dealing with the apps on the screens, he demonstrated competency in engaging with digital texts for meaning-making using viewing skills.

In the second app, Ben chose “Catch a Star” with subtraction on 1 – 10. The questions were in the form of $x – 5 = 2$ and $5 – x = 2$, the task being to determine the value of $x$. Ben was not able to tackle the $x – 5 = 2$ form, and after two wrong attempts, sought help from his teacher. The teacher showed him a trick to sum the numbers of 5 and 2. He managed to acquire knowledge of the new strategy and practised for the next few similar questions. Through reflection, Ben learned the strategy of “addition to solve subtraction problems” (Peters, Smedt, Torbeyns, Verchaffel, & Ghesquière, 2013). In Ben’s classroom activities with the fourth app, he was observed enacting viewing skills that resulted in his learning the new strategy of recalling number, with no apparent counting with observational learning. However, observations of his attempts in using the first and third apps reveal no evidence of a new strategy having been learned.

**Participant four: Nicole**

Nicole was a shy but obedient student.

The data collection took six months, with another two months required to complete the data analysis. In accordance with convention, the activity system used here comprises multiple components. Sub-activity triangles between the components (see Figure 1) were examined closely. An activity system consists of multiple components. The subject and the outcome of the activity refer to the participant and his/her ability to learn a new strategy. The object refers to the reason the participant engages in the activity; the tools are the digital texts of the apps; the rules are the regulations that govern the activity set by the teacher or/and the rules of the apps; the community refers to the environment and how the participants relate explicitly and implicitly to one another during the activity; and the division of labour refers to how the participants differ in the degree of their engagement in activities.

Table 5 summarises the participants’ successful learning experiences. Investigating the first dimension—subject-tool-object—sheds light on the first research question. It reveals that viewing skills helped the participants in active
knowledge construction as they were enacting their tasks temporally, hence externalising their understanding. With the appropriate collaborative learning skills (see Table 6), the enactment of the viewing skills assisted in internalising the new strategies. This, therefore, answers the first research question.

Table 5. Study of viewing and collaborative learning skills

<table>
<thead>
<tr>
<th>Participant</th>
<th>The tool</th>
<th>App</th>
<th>New strategy</th>
<th>Collaborative learning skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter</td>
<td>Enacting viewing skills with digital texts allowed Peter to externalise his strategy (as it was understood by Mary) and internalise the new strategy as he attempted several times to competently apply it. Applying viewing skills in the second app (the button turned blue, indicating “WELL DONE!”) allowed Peter to externalise his strategy. He made several attempts applying the new strategies until he could solve the questions within the time limit. In the fourth app, Peter observed on screen how others had answered the questions, and with the help of loud-answer suggestions from the group internalised his new strategy.</td>
<td>1st</td>
<td>Counting down from the bigger number</td>
<td>Peer learning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>Counting on from the first number</td>
<td>Seeking solutions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4th</td>
<td>Recalling number facts from memory</td>
<td>Observation</td>
</tr>
<tr>
<td>Mary</td>
<td>Mary internalised a new strategy as she attempted the questions on advice from her teacher. She had also internalised Peter’s strategy when she enacted the questions and when working with the fourth app on the screen in front of the group.</td>
<td>2nd</td>
<td>Decomposition</td>
<td>Scaffolding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4th</td>
<td>Recalling number facts from memory</td>
<td>Observation</td>
</tr>
<tr>
<td>Ben</td>
<td>In the second app, Ben internalised his new strategy after a few attempts and on the advice of the teacher. In the fourth app, he also demonstrated using viewing skills to externalise his understanding on the screen.</td>
<td>2nd</td>
<td>Addition to solve subtraction problems</td>
<td>Reflection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4th</td>
<td>Recalling number facts from memory</td>
<td>Observation</td>
</tr>
<tr>
<td>Nicole</td>
<td>In the first app, Nicole demonstrated enacting viewing skills to internalise a new strategy by imitating Ben. The fourth app also allowed her to enact viewing skills to internalise the new strategy.</td>
<td>1st</td>
<td>Counting on from the first number</td>
<td>Imitation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4th</td>
<td>Recalling number facts from memory</td>
<td>Observation</td>
</tr>
</tbody>
</table>

Table 6. Collaborative skills enacted through interaction with digital texts

<table>
<thead>
<tr>
<th>Collaborative skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skills</td>
</tr>
<tr>
<td>Peer learning</td>
</tr>
<tr>
<td>Embracing scaffolding</td>
</tr>
<tr>
<td>Seeking solutions</td>
</tr>
<tr>
<td>Reflection</td>
</tr>
<tr>
<td>Observational learning</td>
</tr>
<tr>
<td>Imitating</td>
</tr>
</tbody>
</table>

This multiple case study provides a more generalisable result on how digital texts, through viewing and representing skills, merge with the collaboration of the participants iteratively to enable the meaning-making process (see Figure 3). However, as observed in the findings of the four participants, some of the interactions with the apps might not result in the gaining of new knowledge. This indicates that other components must be further investigated.

In the second dimension, subject-community-object, the community setting provides the context in which the individual’s learning interaction (articulated interplay) takes place through the activity (see Table 7). Although the findings of previous studies claim that children’s feelings of isolation and a sense of presence were positively
correlated with the effectiveness of their learning (Rovai & Wighting, 2005), the current research results show that
the community setting does not have an absolute effect in new meaning-making. In the second app, three out of four
of the participants gained new knowledge, but in the case of the third app, none did so, although it should be noted
that both the second and the third apps were individually operated by the participants. Although it was demonstrated
that the chances of learning new strategies were highest when the participants enacted viewing skills in a group, the
presence of other components enabled the participants to individually gain new strategies.

![Figure 3. Viewing skills facilitate collaboration among peers](image)

### Table 7. New strategies acquired relative to their social level

<table>
<thead>
<tr>
<th>Community</th>
<th>Articulated interplay level</th>
<th>The new strategies learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team of two (1&lt;sup&gt;st&lt;/sup&gt; app)</td>
<td>Medium</td>
<td>✓</td>
</tr>
<tr>
<td>Individual (2&lt;sup&gt;nd&lt;/sup&gt; app)</td>
<td>Low</td>
<td>✓</td>
</tr>
<tr>
<td>Individual (3&lt;sup&gt;rd&lt;/sup&gt; app)</td>
<td>Low</td>
<td>✓</td>
</tr>
<tr>
<td>Performed one by one in the group (4&lt;sup&gt;th&lt;/sup&gt; app)</td>
<td>High</td>
<td>✓</td>
</tr>
</tbody>
</table>

*Note. “✓” refers to the indication of a new strategy learned.

In the third dimension, subject-division of labour-object, the individual’s roles within the community differ in terms
of the degree of engagement in the activities. The participants exhibited different personalities: Peter and Mary
preferred to deal with people, although Mary was smarter than Peter. Ben was excellent in inductive reasoning and
Nicole was likely to solve problems in an intuitive trial-and-error manner (Kolb, 1981; McLeod, Carpenter,
McCornack, & Skvarcius, 1978). However, both Ben and Nicole were shy, with neither being assertive when they
encountered problems in answering their questions (see Table 8). Nevertheless, it was noticed that Peter learned
three new strategies from the four apps, indicating that his assertive personality helped him acquire more new
knowledge (see Table 8).

### Table 8. Different learning styles versus the new knowledge learned

<table>
<thead>
<tr>
<th>Learning style/ Performance</th>
<th>Result is above average</th>
<th>Result is average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assertive</td>
<td>Mary</td>
<td>Peter</td>
</tr>
<tr>
<td></td>
<td>✓ 0 0 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Shy and not outspoken</td>
<td>Ben</td>
<td>Nicole</td>
</tr>
<tr>
<td></td>
<td>0 0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note. “O”” refers to the frequency of a new strategy learned.

In the fourth dimension, subject-rule-object, all participants were observed to acquire new strategies in the course of
attempting to adhere to instructional requirements affecting scores and/or speed limit when completing tasks.
Intellectual engagement with outcomes and standards were the focuses of participants engaged in answering the
questions. The rules were different in different apps (see Table 9). The results demonstrated that the rules of the
games were important in helping the participants to learn the new strategies. The third app was not designed with any
rules that might help the participants to learn, although the app might have the potential to guide participants to new
knowledge.

386
Table 9. The standard of different apps

<table>
<thead>
<tr>
<th>The apps</th>
<th>The standard</th>
<th>The new strategy adopted by the four participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition &amp; Subtraction for Kids</td>
<td>Result score</td>
<td>oo</td>
</tr>
<tr>
<td>Math Adventure</td>
<td>Time limit and result score</td>
<td>ooooo</td>
</tr>
<tr>
<td>Juicy Math</td>
<td>Time limit</td>
<td>oooo</td>
</tr>
<tr>
<td>Worm Jump</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. “o” refers to the frequency of a new strategy learned.

The analysis of the four dimensions is summarised as follows.

- Viewing and collaborative skills required in interaction with the digital texts enabled the externalisation and internalisation of the understandings, resulting in new meaning-making (see Figure 3). However, the interplay and the sequence of the different dimensions of the activities should be considered to establish a more robust framework for designing the activities (Granberg & Olsson, 2015).

- Participants were immersed in non-routine problems during their enactment with the apps under certain rules and as per the task instructions (see Table 5).

- The differences in personalities of the participants might significantly affect their capacity to learn new strategies (Table 8 refers). Participants’ personalities started to affect ability to learn a new strategy only after their encountering a non-routine problem.

- When participants encounter a non-routine problem, community settings have an effect on the opportunity to learn a new strategy. The type of rules embedded in the game designs might play important roles in this regard; the rules were in place before the participant encountered the non-routine problem (see Table 9).

- The new strategy was learned through applying it to the apps only when the non-routine problem had been solved using collaborative learning skills derived from the peer community.

**Discussions**

Enactment of viewing skills with apps allows the participants to encounter non-routine problems. The ability to enact the apps and solve a problem with a new strategy helps the participants to internalise new knowledge. The findings of research question two are, indeed, analogous to a jigsaw puzzle. Through application of the activity theory, the four dimensions of analysis served as fundamental pieces of the puzzle. Once these pieces were fitted together, the answer to research question two was answered in totality. The summarised framework is presented in Figure 4.

As indicated in Figure 4, the participant starts with a prior knowledge level. In enacting the viewing skills there are four possibilities:

- The rules of the app are followed and the problem is solved with prior knowledge.
- The rules of the app are not followed, and the problem is solved with prior knowledge.
- The rules of the app are followed, and the non-routine problem is encountered but not solved.
- The rules of the app are followed, the non-routine problem is encountered, a new strategy is learned, and the problem is solved by applying the new application.

Through the enactment of viewing skills, a non-routine problem was encountered. The process of negotiation in seeking the new strategy happened at the collaboration level. The interactions between the participants and the digital artefacts transcended their knowledge levels through a positive meaning-making path; this might not have been possible if the contextual setting had not been aligned along the path (see Figure 4). The iterative negotiation between these two skills (viewing and collaborative) led to the internalisation of the new strategy. The framework is an important consideration in designing digital and non-digital activities for young children that lead to self-directed learning.

The current research has several limitations. Although the study focused on the implementation of classroom mathematics learning through activities designed accordingly, the investigation was confined to the topics of addition and subtraction. Thus, findings may not be generalizable to all mathematics learning in the classroom. Future research could focus on more topics with a view to developing the current framework.
Further, the kindergarten encountered difficulties in selecting suitable apps for learning purposes. The apps used in the classroom were limited by the participants’ age-related cognitive abilities, which determined the level of written texts used and the aspects of virtual reality of the app content. Doubt must be acknowledged as to the general validity and availability of apps suitable for young children’s classroom learning.

Figure 4. The framework for enacting viewing skills to acquire new knowledge

Conclusions

This paper synthesises relevant literature on collaborative learning and draws on viewing skills to make a case for integrating digital devices in classroom mathematics learning. Movements are currently underway to reform the practices of mathematics problem solving by focusing on the flexible use of appropriate strategies, rather than standard school-taught approaches (Peters et al., 2013; Verschaffel, Greer, & De Corte, 2007). Learning takes place effectively when learners are in control of the learning activity, able to assess and experiment with their ideas in the course of pursuing results and enquire by working with people in seeking new knowledge before planning for new actions (Ravenscroft, 2000).

The need for learners to externalise understanding is central to the activities. All parties focus on a common external representation of a subject that allows them to identify and discuss the topics (Laurillard, 2002; Pask, 1976). A pedagogical implication arises from the current research. The findings here imply that conventional classroom learning theories regarding young children’s learning of addition and subtraction need to be reconceptualised, and flexible use of strategies in classroom activities with digital technology are promoted and adopted to facilitate collaborative creative reasoning within the proposed framework (see Figure 4). The current multiple case study also provides a framework for the integration of digital texts into activity-based classroom mathematics learning, along with specific recommendations and implications that emerged from the research findings. Activities using digital technologies allow participants to enact autonomous tasks to construct their own meanings.
References


Erratum to - Volume 17, Issue 4, pp. 229-241

Table 1 on page 231 had a mistake in defining the study group. The research was conducted in two groups instead of three.

The corrected Table 1 is as follows:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Pre-test</th>
<th>Experimental Process</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group I</td>
<td>O_{1.1}</td>
<td>X_1</td>
<td>O_{1.2}</td>
</tr>
<tr>
<td>Control Group</td>
<td>O_{2.1}</td>
<td>X_2</td>
<td>O_{2.2}</td>
</tr>
</tbody>
</table>