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 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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.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:

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

The journal is a quarterly journal published in January, April, July and October.

Copyright:

Copyright © 2008, Educational Technology & Society. All rights reserved.

ISSN 1436-4522 (online) and 1176-3647 (print). The articles in this issue of the Journal of Educational Technology & Society are available under Creative Commons CC-BY-NDNC 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Journal Editors at ets-editors@ifets.info.

Editors

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

Editors’ Advisors

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

Editorial Assistant

Barbara Adamski, Athabasca University, Canada; Stelios Sergis, University of Piraeus & Centre for Research and Technology Hellas, Greece.

Associate editors

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

Assistant Editors

Yuan-Hsuan (Karen) Lee, National Chiao Tung University, Taiwan; Wei-Chieh Fang, National Sun Yat-sen University, Taiwan; Sie Wai (Sylvia) Chew, National Sun Yat-sen University, Taiwan.

Advisory board

Ignacio Alonso, 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; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Ruddy Leolouche, 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; Ricchiros Mizonoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan; Yu-Mei Wang, University of Alabama at Birmingham, USA.

Technical Manager

Panagiotis Zervas, University of Piraeus & Centre for Research and Technology Hellas, Greece.
Executive peer-reviewers
http://www.ifets.info/

Supporting Organizations
Centre for Research and Technology Hellas, Greece
Athabasca University, Canada

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

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

Abstracting and Indexing

Guidelines for authors
Submissions are invited in the following categories:

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

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

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

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

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

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

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

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

**Guest Editorial**

Managing Cognitive Load in Technology-Based Learning Environments
Slava Kalyuga and Tzu-Chien Liu

1–8

**Special Issue Articles**

**Cognitive load aspects of technology**

Engaging or Distracting: Children’s Tablet Computer Use in Education
Rhonda N. McEwen and Adam K. Dubé

9–23

Do Learner Characteristics Moderate the Seductive-Details-Effect? A Cognitive-Load-Study Using Eye-Tracking
Babette Park, Andreas Korbach and Roland Brünken

24–36

Gender Effects When Learning Manipulative Tasks From Instructional Animations and Static Presentations
Mona Wong, Juan C. Castro-Alonso, Paul Ayres and Fred Paas

37–52

The Impact of Supported and Annotated Mobile Learning on Achievement and Cognitive Load
Rustam Shadiev, Wu-Yuin Hwang, Yueh-Min Huang and Tzu-Yu Liu

53–69

Effects of Computer-Based Visual Representation on Mathematics Learning and Cognitive Load
Hsin I., Yung and Fred Paas

70–77

**Management of cognitive load for effective use of technology**

Designing Effective Video-Based Modeling Examples Using Gaze and Gesture Cues
Kim Ouwehand, Tamara van Gog and Fred Paas

78–88

Computer-Based Learning of Geometry from Integrated and Split-Attention Worked Examples: The Power of Self-Management
Sharon Tindall-Ford, Shirley Agostinho, Sahar Bokosmaty, Fred Paas and Paul Chandler

89–99

Learning from Concept Mapping and Hypertext: An Eye Tracking Study
Franck Amadieu, Ladislao Salmerón, Julien Cegarra, Pierre-Vincent Paubel, Julie Lemarié and Aline Chevalier

100–112

Interactions Between Levels of Instructional Detail and Expertise When Learning with Computer Simulations
Yuling Hsu, Yuan Gao, Tzu-Chien Liu and John Sweller

113–127

Using the Multi-Display Teaching System to Lower Cognitive Load
Tsung-Sheng Cheng, Yu-Chun Lu and Chu-Sing Yang

128–140

An Analytics-Based Approach to Managing Cognitive Load by Using Log Data of Learning Management Systems and Footprints of Social Media
Cheng-Huang Yen, I-Chuan Chen, Su-Chun Lai and Yea-Ru Chuang

141–158

Interactivity of Question Prompts and Feedback on Secondary Students’ Science Knowledge Acquisition and Cognitive Load
Kun Huang, Ching-Huei Chen, Wen-Shiuan Wu and Wei-Yu Chen

159–171

**New trends and directions in cognitive load research on technology-based learning environments**

Metacognitive Load – Useful, or Extraneous Concept? Metacognitive and Self-Regulatory Demands in Computer-Based Learning
Rolf Schwonke

172–184

The Effects of Rapid Assessments and Adaptive Restudy Prompts in Multimedia Learning
Alexander Renkl, Irene T. Skuballa, Rolf Schwonke, Nora Harr and Jasmin Leber

185–198

ISSN 1436-4522 (online) and 1176-3647 (print). The articles in this issue of the Journal of Educational Technology & Society are available under Creative Commons CC-BY-NDNC 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Journal Editors at ets-editors@ifets.info.
Using Cognitive Load Theory to Tailor Instruction to Levels of Accounting Students’ Expertise

Paul Blayney, Slava Kalyuga and John Sweller

The Effects of Social Cue Principles on Cognitive Load, Situational Interest, Motivation, and Achievement in Pedagogical Agent Multimedia Learning

Sanghoon Park

The Beast of Aggregating Cognitive Load Measures in Technology-Based Learning

Jimmie Leppink and Jeroen J. G. van Merriënboer

Full Length Articles

Characterization of Educational Resources in e-Learning Systems Using an Educational Metadata Profile

Georgia Solomou, Christos Poirrakes and Achilles Kameas

From Motivation to Engagement: The Role of Effort Regulation of Virtual High School Students in Mathematics Courses

ChanMin Kim, Seung Won Park, Joe Cozart and Hyewon Lee

Automatic Scaffolding and Measurement of Concept Mapping for EFL Students to Write Summaries

Yu-Fen Yang

Comparison of Collaboration and Performance in Groups of Learners Assembled Randomly or Based on Learners’ Topic Preferences

Karina L. Cela, Miguel Ángel Sicilia and Salvador Sánchez

An Integrated Contextual and Web-based Issue Quest Approach to Improving Students’ Learning Achievements, Attitudes and Critical Thinking

Han-Yu Sung, Gwo-Jen Hwang and Hong-Sheng Chang

Pedagogically-Driven Ontology Network for Conceptualizing the e-Learning Assessment Domain

Lucila Romero, Matthew North, Milagros Gutiérrez and Laura Caliusco

Effects of Annotations and Homework on Learning Achievement: An Empirical Study of Scratch Programming Pedagogy

Addison Y. S. Su, Chester S. J. Huang, Stephen J. H. Yang, T. J. Ding and Y. Z. Hsieh

Learning Style and Task Performance in Synchronous Computer-Mediated Communication: A Case Study of Iranian EFL Learners

Mohsen Hedayati and Elham Mohammadi Foomani

Has Research on Collaborative Learning Technologies Addressed Massiveness? A Literature Review

Kalpani Manathunga and Davinia Hernández-Leo

Facebook Groups as an Academic Teaching Aid: Case Study and Recommendations for Educators

Eli Miron and Gilad Ravid

Designing a Resource Evolution Support System for Open Knowledge Communities

Xianmin Yang and Shengquan Yu

Impacts of Pedagogical Agent Gender in an Accessible Learning Environment

Noah L. Schroeder and Olusola O. Adesope

Value-Added Results for Public Virtual Schools in California

Richard Ford and Kerry Rice

Instructional Design and Professional Informal Learning: Practices, Tensions, and Ironies

Stephen C. Yanchar and Melissa N. Hawkley

A Review of Emotion Regulation in Intelligent Tutoring Systems

Mehdi Malekzadeh, Muntaq Begum Mustafa and Adel Lahsasna

Learner-Centered Blogging: A Preliminary Investigation of EFL Student Writers’ Experience

Ming Huei Lin

Exploring Effects of Multi-Touch Tabletop on Collaborative Fraction Learning and the Relationship of Learning Behavior and Interaction with Learning Achievement

Wu-Yuin Hwang, Rastam Shadiev, Chi-Wei Tseng and Yueh-Min Huang
A Robot-Partner for Preschool Children Learning English Using Socio-Cognitive Conflict
Elvis Mazzoni and Martina Benvenuti

Language Learning in Virtual Reality Environments: Past, Present, and Future
Tsun-Ju Lin and Yu-Ju Lan

Acceptability and Satisfaction of an ICT-based Training for University Teachers
Rocio Herrero, Juana Bretón-López, Luis Farfallini, Soledad Quero, Ignacio Miralles, Rosa Baños and Cristina Botella

Wiki-Mediated Activities in Higher Education: Evidence-Based Analysis of Learning Effectiveness Across Three Studies
Panagiota Altanopoulou, Nikolaos Tselios, Christos Katsanos, Maria Georgoutsou & Maria-Antonia Panagiotaki

Book Reviews

Computer Games for Learning: An Evidence-Based Approach (Author: Richard E. Mayer)
Reviewer: Colin Pinnell

Embracing Social Media: A Practical Guide to Manage Risk and Leverage Opportunity (Author: Kristin Magette)
Reviewer: Jon Dron
Guest Editorial: Managing Cognitive Load in Technology-Based Learning Environments

Slava Kalyuga¹ and Tzu-Chien Liu²*

¹School of Education, University of New South Wales, Sydney, Australia // ²Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan // s.kalyuga@unsw.edu.au // tzuchien@ntnu.edu.tw

*Corresponding author

ABSTRACT

Cognitive load theory is an instructional theory that uses our knowledge of human cognitive architecture, especially processing limitations of working memory, to enhance effectiveness of instructional design. This paper reviews main assumptions and principles of cognitive load theory and discusses their application to technology-based learning environments. The paper concludes with a brief introduction to the structure and content of this Special Issue.

Keywords

Cognitive load theory, Technology-based learning

Introduction

Technology-based learning environments offer a wide range of educational opportunities that could not be achieved in traditional face-to-face forms of learning and instruction. For example, learners could be flexible in using their time, have increased levels of control over the content and pace of presentation, and have information independently delivered to a convenient place. However, as with all technology applications, the use of technology by itself may not guarantee sufficient benefits for learning to warrant investing into the development of such applications. In particular, technology applications may pose additional processing demands on learners' cognitive resources, which may negatively affect the construction of new knowledge. In general, empirical research into technology-based learning has produced mixed findings that support the above concern and make clear that the analyses of associated cognitive processes and structures are essential for enhancing the effectiveness of technology-based learning environments. The cognitive load imposed on learners in technology-based environments is the main focus of this special issue, and cognitive load theory is the main theoretical framework for all the contributing papers.

Over the past decades, cognitive load theory has advanced rapidly and has been applied to enhance the effectiveness of learning and instruction in various subject areas (see Sweller, Ayres, & Kalyuga, 2011, for the most recent overview of the theory). In addition to a great number of individual papers reporting theoretical and empirical studies within a cognitive load framework published every year in multiple journals in the areas of educational psychology, learning and instruction, and the related fields, there have been several special edited issues intended to guide the new directions for the development of cognitive load theory and also document the contributions of cognitive load theory to instructional design across different disciplines, for example, the special issues published in Educational Psychologist (2003), Educational Technology Research and Development (2005), Learning and Instruction (2009), Educational Psychology Review (2007, 2010, 2015), Instructional Science (2010). Although a number of recent papers have considered cognitive load factors influencing learning in technology-based environments (e.g., Liu, Lin, Tsai, & Paas, 2012; Kalyuga, 2012), this is essentially the first special issue specifically devoted exclusively to discussing the innovative technology-supported learning from a cognitive load perspective.

Cognitive load theory

Cognitive load theory (CLT) is an instructional theory developed to coordinate instructional design and learning procedures with human cognitive architecture (Sweller, Van Merrienboer, & Paas, 1998; Sweller et al., 2011). In the most general form, this cognitive architecture consists of two main components: (1) working memory as our main processor of information with a very limited capacity and duration when dealing with novel, unorganized information, and (2) effectively unlimited long-term memory storing cognitive schemas (knowledge structures we
use to categorize information for intended use) that vary in their degree of complexity and automation. These two components of human cognitive architecture are closely interrelated with each other.

On the one hand, new knowledge structures that are consciously constructed in working memory become a part of the knowledge base in long-term memory. The conscious construction in working memory is the main way of the acquisition of new knowledge in educational settings. It is different from the way we acquire common skills such as speaking and listening in everyday native language, basic social interaction skills, etc. We are believed to be genetically predisposed to acquire such skills in an unconscious, intuitive way by being immersed in the corresponding social environments (Geary, 2007). On the other hand, organised knowledge structures in long-term memory, when activated and brought into working memory, allow reducing the processing load in working memory by encapsulating many familiar elements of information into larger units (schemas) that are then treated as single elements in working memory. In this way, the effective capacity of working memory could be significantly increased; the upper limits of this capacity depend only on the available relevant knowledge base. For example, an unfamiliar written foreign word would normally be processed as a sequence of related letters or symbols that would constitute the elements to be processed in working memory. However, when it is learned and become part of the knowledge base, the whole word would be treated as a single element. Thus the actual capacity of working memory is always dependent on the available relevant content of long-term memory.

Cognitive load could be defined as working memory resources required for completing a learning task or activity by learners with a specific level of prior knowledge. It is believed that the magnitude of cognitive load is determined by the degree of element interactivity – interconnectedness between the related elements of information that need to be processed simultaneously in working memory. Higher levels of cognitive load are usually caused by learning tasks with high degrees of element interactivity (e.g., learning complex grammar rules of a foreign language). If individual elements of information could be processed separately without the need to attend to other related elements, such materials would not cause any cognitive load issues (e.g., learning vocabulary items in a foreign language).

There are two major types of cognitive load distinguished in cognitive load theory. Intrinsic cognitive load is the load that is relevant to achieving the intended learning goals. It depends on the degree of interactivity between essential elements of information and on the level of learner prior knowledge in long-term memory. Extraneous cognitive load is the load that is irrelevant to learning and is present only because instruction is designed in a way that requires learners to engage in cognitive processes and activities that are not actually required for acquisition of intended schemas. For example, when two related sources of information – such as a diagram with explanatory text - that need to be processed at the same time to understand the instructional message are separated in space (e.g., placed at the opposite sides of a screen or in different windows) or in time (e.g., presented sequentially rather than simultaneously), maintaining some elements of these sources while searching for their co-referents may consume working memory resources that would be unavailable for meaningful learning. Another typical situations that may cause extraneous cognitive load are presenting redundant sources of information that duplicate the already available and sufficient for learning source (e.g., an explanatory text in addition to a self-explanatory diagram) or presenting information that is redundant to more knowledgeable learners (while still essential for novices). In the above cases, processing redundant sources of information could unnecessarily consume working memory resources that would be unavailable for further learning.

In addition to the above dimension of cognitive load (as required resources for a task), cognitive load theory also considers the dimension of working memory resources actually allocated by the learner to the learning task. It is an important concept, since if no resources are actually allocated to the learning task, nothing would likely be learned (irrespective of what the required resources or cognitive load). The critical for learning part of this dimension is the concept of germane working memory resources (or, as it still commonly called, germane cognitive load) – the resources actually allocated by the learner to processing essential elements of instruction or, in other words, to dealing with intrinsic, essential cognitive load. The level of actually allocated resources (most importantly, germane resources) usually depends on learner motivation and engagement with the learning task. These factors could be outside of control of traditional instructional design techniques developed by cognitive load theory and might depend on professional skills of teachers. Still, the presence of cognitive overload by itself could be considered as one of the factors that might negatively influence learner motivation to learn.

Cognitive load theory has generated many instructional techniques (usually called cognitive load effects) to reduce learner cognitive load – especially extraneous load. For example, one of the best-known and well-established effects
is the worked example effect, according to which providing explicit guidance in the form of worked-out solution steps could result in better learning for novices in comparison with solving an equivalent number of problems due to reducing the extraneous load caused by unguided search processes. The split-attention effect suggests integrating related sources of information (e.g., embedding some of them within others or synchronizing them in time) that could not be understood in isolation instead of providing them in separated, split source formats that could generate extraneous load. The redundancy effect suggests the elimination of any sources of information that re-describe in a different form a source of information that could be understood on its own. Processing any redundant information would cause unnecessary extraneous cognitive load. According to the modality effect, textual information accompanying pictures or animations should be presented in an auditory rather than visual form in order to engage both available channels of working memory – auditory and visual - instead of overloading one of them. According to the audio-visual (multimedia) redundancy effect, the same verbal information should not be presented simultaneously in auditory and visual modalities to avoid unnecessary cognitive load imposed by the need to correlate these sources of information. All of these effects could be applied in contemporary technology-based learning environments.

Cognitive load factors in technology-based learning

Modern technology-based learning environments use the whole range of technological innovations to improve the quality of learner experiences and activities, and in this way enhance their learning outcomes. Such innovations include, among others, using spoken explanations, different forms of dynamic visualizations (e.g., animations and simulations), connecting such multimedia components into hypermedia learning environments by sophisticated networks of hyperlinks, using virtual learning environments and sophisticated online collaboration tools. Still, many decades of continuing attempts to improve education by introducing new technological innovations into classrooms and other educational settings have taught us one important lesson: it is not technology itself that matters but how it is used, whether it would elicit learner activities and cognitive processes that could lead to constructing better and more durable knowledge. This understanding would inevitably lead us to human cognitive architecture and its major features as critical factors that need to be considered in order to assure instructional effectiveness of any technological innovation. Ignoring these factors and cognitive characteristics of learners may result in continuing disappointments.

For example, the two most basic and widespread technological innovations of recent decades that have inspired great expectations of better learning – transforming written text into spoken narrations and transforming static pictures into dynamic visualizations such as animations or simulations – produced rather mixed results in reality. Closer look at these technological means shows that both transform permanent information into a transient form. Both spoken words and dynamic visuals could disappear from learner auditory and visual sensory registers very rapidly and before they are integrated with the following information to achieve understanding. To deal with such transience, the learner must maintain these fleeting elements of information in a temporary storage – which is working memory – before the integration happens. Such processes may easily overwhelm working memory (especially for novice learners whose working memory is most severely limited) and inhibit learning instead of improving it. In fact, some reported failures in observing two well established cognitive load effect in multimedia learning that support the use of spoken instead of written words – the modality and verbal redundancy effects - as well as failures to consistently demonstrate benefits of animations over static graphics, have been recently explained by the effects of transient information in the corresponding conditions (the transient information effect in cognitive load theory – see Leahy & Sweller, 2011; Sweller et al., 2011). Learner-controlled pacing of presentations supplemented with segmenting materials into smaller sections cold help in dealing with negative consequences of transient information (Singh, Marcus, & Ayres, 2012).

High levels of cognitive load in high-tech learning environments could be also generated by the need for learners to distribute their attention between several related sources and modes of information presentation that should be processed concurrently in working memory, or to deal simultaneously with many variables, often in conditions of uncertainty and non-linear relationships (e.g., in hypertext and hypermedia learning environments). Such environments may easily create split-attention situations. It could be as simple and common situations as feedback on learner performance that shows at a distance from the task or in a separate window that covers the original learner answers. Physically integrating the related sources of information (or synchronizing them in time) could reduce the corresponding extraneous load. Using appropriate visual cues may also reduce split attention by directing learner
attention to essential elements of information. In more complex situations, developing effective technical means for monitoring and tracking concurrent changes in dynamic information presentations could reduce the levels of associated cognitive load.

Surprisingly, some instructional procedures and techniques used in technology-based learning environments to manage cognitive load may potentially contribute to both reduction and increase in extraneous load, thus either enhancing or hindering learning. For example, technology-based learning environments often involve high levels of learner control. Learners may need to make their own decisions on sequencing the content, pacing the presentation, or using the available instructional support means and specific forms of such support. Such cognitive processes could require additional working memory resources, thus potentially increasing levels of cognitive load. If advanced learners with higher levels of prior knowledge could handle this load, it could inhibit learning for less experienced learners, and suitable instructional techniques should be used to prevent this. On the other hand, techniques for enhancing learner control over the pace of presentation (e.g., using navigation tools) could reduce extraneous cognitive load caused by information transiency.

Dynamic visualizations (such as animations, simulations, dynamic concept maps) and hypermedia environments may be effective in representing new concepts in dynamic formats. However, when using such learning tools, often supplemented with sophisticated navigational features, learners may overlook the learning goals and instead, search for exciting and entertaining material. Managing the complexity of such environments and providing the learners with the learning tasks that are carefully tailored to their cognitive characteristics could facilitate essential processing and reduce irrelevant activities.

Thus, specific conditions of effectiveness of different instructional methods in technology-based learning environments, especially the role of learner prior knowledge, may need to be considered in each particular case instead of following a universal fixed set of principles or guidelines. The general cognitive load principle in application to the design of such learning environments could be expressed as enhancing cognitive processes that are essential to knowledge construction without triggering cognitive processes that increase extraneous cognitive load. Cognitive load theory can provide an appropriate framework for evaluating the sustainability of learner cognitive processes involved in technology-based learning and developing effective instructional approaches for enhancing instructional efficiency of such learning environments by optimizing their cognitive load conditions. With this theory, the technology-based learning environments could be better matched to the nature of human cognition.

The cognitive load aspects of learning in technology-based learning environments are the main focus of contributions to this Special Issue. This Issue goes further than just replicating some of the already investigated and established instructional methods in new high-tech learning environments. The papers in this issue also report some innovative results related to such hot educational research issues as enhancing learner motivation and positive attitudes, levels of their self-regulation skills and metacognitive strategies, taking into account learner gender differences, developing learner-tailored adaptive environments, implementing more powerful diagnostic tools such as eye-tracking techniques to uncover the important patterns in cognitive processes and their relations with learner cognitive characteristics to provide more customized support functionality, and others.

The general structure of this special issue

This special issue offers insights into the current and future trends and research directions in managing cognitive load in technology-based learning. It was intended to focus especially on such important themes as cognitive load implications for multimedia learning, mobile learning, simulation- and game-based learning, virtual reality-based learning, technology enhanced language learning, and the other areas relevant to technology-based learning. The selected contributions reflect one of the following three main tracks:

Cognitive load aspects of technology used in the educational contexts. This track relates to the exploration and evaluation of cognitive load resulted from learning in technology-based environments. It not only refers to the effective, intrinsic cognitive load induced by such environments, but also to the unnecessary cognitive load imposed when using inappropriate technology, or using technology in an inappropriate way. Special interest is posed in ongoing projects investigating the appropriateness of using technologies in education from the perspective of cognitive load theory.
Management of cognitive load for effective use of technology in the educational contexts. This track includes studies investigating specific strategies for managing cognitive load in using technology with the aim of promoting effective cognitive load and reducing extraneous cognitive load. Such studies may provide the basis for new theoretical approaches to the design of technology-based learning environments in specific areas of e-learning, such as computer-based learning environments, simulation-based learning environments, mobile device assisted learning environments, virtual classroom learning environments, etc.

New trends and directions in cognitive load research on technology-based learning environments. Although cognitive issues have usually been neglected during the rapid development of technology designed to enhance education, cognitive load theory has identified these issues as significant factors influencing the instructional effectiveness of technology. Research studies in technology-based environments based on cognitive load approaches continue to identify new issues and cognitive load aspects that should direct the researchers and educators in justifying the effectiveness of new technological developments and applications in education. In addition new theoretical perspectives and methodological approaches have been developed that could be highly relevant to this area of research.

Brief overview of contributions to the special issue

According to the above structure, the following papers contributed to the corresponding main tracks.

Cognitive load aspects of technology

In their paper “Engaging or distracting: children’s tablet computer use in education,” Rhonda McEwen and Adam Dube investigated the relationship between a user’s engagement with tablet computers and user’s cognitive load. Eye tracker was applied as an important instrument in data collection. The results demonstrated that in spite of the children’s cognitive abilities, simpler applications were better able to direct their attention to the essential, intrinsic content.

The paper “Do learner characteristics moderate the seductive-details-effect? A cognitive-load-study using eye-tracking” by Babette Park, Andreas Korbach, and Roland Brünken examined whether the seductive details effect in multimedia biology learning is moderated by spatial ability and prior knowledge. The analysis of perceptual processing measured by eye-tracking and learning performance showed the overall detrimental effect of seductive details, particularly for learners with low spatial ability and low prior knowledge. Learners with high spatial ability or high prior knowledge could presumably compensate or even profit from seductive details.

The paper “Gender effects when learning manipulative tasks from instructional animations and static presentations” by Mona Wong, Juan C. Castro-Alonso, Paul Ayres, and Fred Paas explored the effect of gender on the effectiveness of static and animated presentations when learning a manipulative task using two different learning platforms - real Lego bricks (physical environment) in Experiment 1 and computerized images of the bricks (virtual environment) in Experiment 2. Gender vs presentation format interaction patterns were found in both experiments. Females performed better than males when learning manipulative tasks with the instructional animations, however no gender differences were found for the static presentations.

“The impact of supported and annotated mobile learning on achievement and cognitive load” by Rustam Shadiev, Wu-Yuin Hwang, Yueh-Min Huang and Tzu-Yu Liu investigated the effects of using a mobile learning environment with authentic support on learning achievement and cognitive load when learning English as a foreign language. The results demonstrated that using the tablet learning system resulted in better post-test performance and caused less cognitive load than when learning without technological support.

In their paper “Effects of computer-based visual representation on mathematics learning and cognitive load,” Hsin I. Yung and Fred Paas demonstrated that an experimental condition with visual representations in mathematics for primary school students resulted in higher learning performance and lower cognitive load than a similar condition without visual representations.
Management of cognitive load for effective use of technology

The paper “Designing effective video-based modeling examples using gaze and gesture cues” by Kim Ouwehand, Tamara van Gog, and Fred Paas investigated whether gaze and gesture cues would improve the distribution of visual attention when studying a video-based modeling example. The eye movement data revealed that when learning from video, students in both gesture + gaze cue condition and no cue condition attend more at the female model than the task area she referred to. The significant interaction between object of attention and condition showed the students in gesture + gaze cue condition had a smaller attentional bias towards the model compared to the task relevant areas than the students in no cue condition. In other words, providing cues, gestures in particular seem effective in redirecting learners’ attention from the model to the task areas the model referred to.

“Computer-based learning of geometry from integrated and split-attention worked examples: The power of self-management” by Sharon-Tindall-Ford, Shirley Agostinho, Sahar Bokosmaty, Fred Paas, and Paul Chandler evaluated the effects of learner self-managing split-attention in worked examples by integrating spatially separated text and diagrammatic information using online tools (e.g., by moving text to associated parts of a diagram) as compared to studying instructor-managed integrated examples in geometry. The results indicated that learners who self-managed split-attention in worked examples performed better on post-test than learners who studied split-attention examples, and performed as well as learners who studied instructor-managed integrated examples.

In the paper “Learning from concept mapping and hypertext: an eye tracking study,” Franck Amadieu, Ladislao Salmerón, Julien Cegarra, Pierre-Vincent Paubel, Julie Lemarié, and Aline Chevalier investigated the effects of prior domain knowledge and learning sequences on learning with concept mapping and hypertext. Participants either made a concept map in a first step and then read the hypertext’s contents combined with concept mapping (high activating condition), or they read the hypertext’s contents first and then made a concept map and re-read the hypertext’s contents (low activating condition). The results confirmed the hypothesis that the low activating condition fostered better learning of relations between the concepts than the high activating condition, regardless of the level of prior knowledge. However, concept mapping behaviors and eye movement data showed that prior knowledge reduced disorientation, improved navigation coherence, and supported better elaboration of semantic relations between the concepts before reading the texts.

The study reported in the paper “Interactions between levels of instructional detail and expertise when learning with computer simulations” by Yuling Hsu, Yuan Gao, Tzu-Chien Liu, and John Sweller used cognitive load theory to investigate the effect of different levels of instructional detail and expertise in a simulation-based environment on learning about concepts of correlation. Results were consistent with the expertise reversal effect, indicating that higher levels of instructional detail benefited learning for lower-expertise learners, whereas lower levels of detail facilitated learning for higher-expertise learners. The authors concluded that the level of instructional guidance needed to match learners’ levels of expertise.

In “Using the multi-display teaching system to lower cognitive load” by Tsung-Sheng Cheng, Yu-Chun Lu, and Chu-Sing Yang, cognitive load principles for the design of multimedia teaching materials were used to develop a multi-display teaching system. This multi-display teaching system was compared to a conventional one-display teaching system. The results revealed that the multi-display teaching system imposed a lower load and resulted in higher learning outcomes.

In the study reported in “An analytics-based approach to managing cognitive load by using log data of learning management systems and footprints of social media” by Cheng-Huang Yen, I-Chuan Chen, Su-Chun Lai, and Yea-Ru Chuang, a technology-based learning environment was used to test the effectiveness of an analytics based approach based on log data of a learning management system and footprints of social media to manage learners’ cognitive load in online courses.

In “Interactivity of question prompts and feedback on secondary students’ science knowledge acquisition and cognitive load”, Kun Huang, Ching-Huei Chen, Wen-Shian Wu, and Wei-Yu Chen examined the effects of problem-solving question prompts and feedback in web-based learning environment on knowledge acquisition and cognitive load. The significant interaction between question prompts and feedback showed that problem-solving prompts combined with corrective feedback can lead to improved knowledge acquisition and reduced cognitive load.
New trends and directions in cognitive load research on technology-based learning environments

In his theoretical paper “Metacognitive load – useful, or extraneous concept? Metacognitive and self-regulatory demands in computer-based learning,” Rolf Schwonke argues that cognitive load theory may profit from considering new types of demands -metacognitive and self-regulation demands - as sources of cognitive load. The analysis of empirical studies showed that computer-based learning environments might pose a variety of cognitive, metacognitive and self-regulatory demands on learners, and most learners might not be able to regulate their learning.

The paper “The effects of rapid assessments and adaptive restudy prompts in multimedia learning” by Alexander Renkl, Irene Skuballa, Rolf Schwonke, Nora Harr, and Jasmin Leber investigated the effects of using embedded rapid diagnostic assessment tasks (designed to be an instrument for building adaptive learning environments) and different adaptive restudy prompts in multimedia learning. The study established that these assessment tasks did not per se influence the measured variable by having positive effects on learning. This is an important result that may justify the use of rapid assessment methods for building adaptive learning environments. It was also demonstrated that unspecific prompts that focus on the corresponding whole knowledge sub-area enhanced learning outcomes more than specific prompts that focus on a very specific piece of knowledge (even though both types of prompts improved learning).

The theme of technology-based adaptive learning environments was further discussed by Paul Blayney, Slava Kalyuga and John Sweller (“Using cognitive load theory to tailor instruction to levels of accounting students’ expertise”) who reported the results of two experiments conducted to investigate the expertise reversal effect based on the isolated-interactive elements instructional procedure and an application of this effect to improve instruction in the domain of accountancy. Experiment 1 found a significant interaction (expertise reversal effect) between isolated-interacting elements effect and levels of expertise on solving complex accounting problems. For more knowledgeable learners, the interactive form of instruction resulted in better learning than the isolated form; but this pattern of results was not obtained for novices. Based on the findings of Experiment 1, Experiment 2 demonstrated that learning could be improved through the adaptation of instructional techniques according to the expertise of the individual. Students for whom instructional method was adapted according to their level of expertise performed better than students who were randomly assigned to an instructional method.

Sanghoon Park (“The effects of social cue principles on cognitive load, situational interest, motivation, and achievement in pedagogical agent multimedia learning”) investigated the effects of social cues on learners’ cognitive load, situational interest, achievement, and motivation in pedagogical agent-based multimedia learning. The results revealed that learners’ situational interest and motivation in terms of relevance and confidence were increased with the use of social cues. Moreover, learners’ situational interest and motivation were influenced by the source of narration. The results may indicate new ways of enhancing germane cognitive resources by increasing learner motivation and engagement with a task – a new and relatively undeveloped issue within a cognitive load framework.

Finally, the paper “The beast of aggregating cognitive load measures in technology-based learning” by Jimmie Leppink and Jeroen J. G. van Merriënboer is concerned with measurement of performance and cognitive load in technology-based learning environments when repeated measurements are taken two or more times on the same variables. The authors argue that the common practice of aggregating scores into a single average score per participant for the subsequent analysis could result in a distorted view of observed effects and miss some potentially important relations of interest. The paper suggests alternative statistical approaches to better account for essential features of the data, thus contributing to the important issue of adequate measurement of cognitive load and learner performance outcomes.

References


Engaging or Distracting: Children’s Tablet Computer Use in Education

Rhonda N. McEwen and Adam K. Dubé

Abstract
Communications studies and psychology offer analytical and methodological tools that when combined have the potential to bring novel perspectives on human interaction with technologies. In this study of children using simple and complex mathematics applications on tablet computers, cognitive load theory is used to answer the question: how successful are tablet computer educational applications at directing children’s attention towards intrinsic and germane content? An eye tracker collected gaze data and cognitive tasks were performed to assess memory and attention. The results show that simple applications are able to direct a child’s attention to intrinsic and germane content, regardless of the child’s cognitive ability. Children assessed as high executive functioning found the germane content of the complex applications helpful whereas children assessed as lower executive functioning did not take advantage of the germane content. Claims that the cognitive structure of the individual is intimately linked to the forms or systems of communication used were partially supported. The research showed that tablet computers and their applications offer a learning experience that appears to be inherently highly interactive—thereby introducing challenges to the cognitive load of children as users.

Keywords
Cognitive load, Eye tracking, Tablet computers, Executive functioning, Child-tablet interaction

Introduction
Tablet computers are being used for educational purposes but there is little examination of how users interactions with these technological objects affect the learning process. Several disciplines have an interest in the relationship between sensory explorations of objects and processes of knowledge creation. Cultural theorists focus on the contribution of societal norms and expectations to epistemic encounters that historically range from 15th century analyses of witchcraft (Classen, 2005), to mid-18th century chemists (Roberts, 2005), and to modern issues involving sensory dis-integration in mental illness (Desjarlais, 2005). Philosophers have a long-standing curiosity in, for example, the roles that sensory perception and belief play in human understanding - spanning physical and metaphysical engagements with everyday objects (Descartes, 1984; Armstrong, 1973; Rorty, 1979; Goldman, 1986). Information and communication scholars consider issues of preservation, material culture and making, memory, and information processing and grapple with the challenges that arise from people’s direct engagement with content (Howarth, 2005; Fisher, Erdelez & McKechnie, 2005). Most notably, cognitive psychologists have established a canon of knowledge contesting and also linking physical experiences, with mental processes and representations. The body of work on cognitive load in particular offers researchers a conceptual framework within which to examine interactions (e.g., Lee, Plass, & Homer, 2006; Paas, Renkl, & Sweller, 2004; van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009). It is the latter two disciplinary approaches that influence the conceptual foundation for this study of how users interact with and make sense of tablet computers. Communications studies and psychology offer analytical and methodological tools, and when combined they have the potential to bring novel perspectives on human interaction with technologies.

Touching, seeing, and cognitive load
Relatively new media devices like smartphones, interactive whiteboards, and tablet computers (e.g., iPads, Innotabs, Android tablets, LeapPads) engage users through touch-interfaces where tactile, visual, and to an optional degree, auditory senses are highly involved in the device-user exchange. In communications studies, scholars like Rowland (2012) theorize that historically technologies have had differential impacts in the defining characteristics of our capacities. Regarding literacy, Botha (1992) posits that “the very cognitive structure of the individual and the formal
patterns of human social relations are intimately linked to the forms or systems of communication [used] sic…” (p. 273). For tablet computers, visual information drives the interaction and likely affects cognitive structure.

In comparison to forms of digital communication available to the general public before sales of tablet computers in 2010, the degree of user-medium interactivity has increased. Yet, questions remain about the extent to which these forms of interaction affect the user, especially questions regarding the nature of the link between communication technologies and potential changes to a user’s “cognitive structure.” Given the broad, rapid adoption of tablet computers in education, it is increasingly important to question how this technology interacts with and affects cognitive structure. This leads to the primary motivation of the present study: to explore the relationship between a user’s engagement with tablet computers and that user’s cognitive load.

Cognitive load results from the short-term information processing activities of the mind when various information elements are being held and manipulated simultaneously (Sweller, 1994). The short-term system for storing and manipulating information is called working memory (Baddeley, 2003) and it is a finite resource that can be overwhelmed (i.e., cognitive overload). To partially overcome the limitations of working memory, information acquired during learning is organized into schemas. A schema is defined as “a cognitive construct that organizes the elements of information according to the manner with which they will be dealt” (Sweller, 1994, p. 296). Cognitive load theory groups schemas created during learning into three types: (a) intrinsic, load that is the inherent to level of difficulty associated with specific content; (b) extraneous, load associated with how information is presented to users; and (c) germane, load generated by the processing, creation, and automation of schemas (Sweller, Van Merriënboer & Paas, 1998). These three types of cognitive load are present to varying levels during all learning tasks and the goal of instruction, within this paradigm, is to best align the learning content to human cognitive architecture. That is, instructional content should be optimized such that cognitive effort is directed towards intrinsic and germane content and away from extraneous content (Paas et al., 2004).

Tablet computers are thought to be effective learning tools because they contain multimedia content that engages users, but it is not known whether users are engaged with the right content. Cognitive load theory has been used to frame how learning generally occurs in multimedia rich environments. Multimedia can be defined narrowly as learning from both words and pictures simultaneously (Mayer, 2001) or more broadly as learning from multiple sensory channels simultaneously (e.g., pictures and audio). When learning using tablet computers, cognitive overload can arise from presenting intrinsic content across both words and pictures simultaneously—such that it cannot be effectively integrated into working memory due to the splitting of attention—or presenting incidental extraneous content in one format that diverts attention away from the intrinsic content presented in another format (Mayer & Moreno, 2003). This overload can be addressed by synchronizing intrinsic content across formats—presenting redundant, reinforcing content that limits the effort required for integration; limiting the competition between extraneous and intrinsic content by reducing their simultaneous presentation (Kaminski & Sloutsky, 2013); or by individualizing content so that it speaks to the cognitive strengths of the user (i.e., visual content for users with larger visual short-term memory; Mayer & Moreno, 2003). Individualizing content is particularly interesting because it highlights how individual differences in users’ cognitive ability influence cognitive overload. It is not known whether the visual content on tablet computers adequately address cognitive overload.

In user-tablet computer interactions the app directs the user’s attention and cognitive activity towards information visually presented on the screen. In investigating tablet computers and the relationship between touching these devices and cognitive load, the contribution of visual information must be considered as it communicates the gestural responses required from the user. This sensory interplay between touching and seeing was also noted by communications theorist McLuhan (2005) when he draws from sculptor Adolf Hildebrand’s insistence in 1893 that “true vision must be much imbued with tangibility.” (p. 43). In addition, keeping with evidence from neuroscience on the interconnectedness of sensory areas of the brain (Sacks, 2005) it is useful to focus attention on what the user looks at when using tablet computers to learn.

Tablet computers and education

There is a longstanding and often controversial tradition of co-opting the use of electronic media technologies for education. From the use of educational radio programs in the classroom in the 1920’s and 1930’s (Atkinson, 1942), to the use of television in elementary school curricula (Cuban, 1986), and the use of computers within formalized
learning settings (Coley, Cradler & Engel, 1997), technologies have continually held the promise of improving the learning process. Tablet computers follow in this trajectory and are increasingly a part of the classroom experience from pre-school to tertiary education (Henderson & Yeow, 2012; Mang & Wardley, 2012, Chatsick, McEwen & Zbitnew, 2013).

When the purchase price of a technology eventually falls within the reach of the middle-class, these devices are increasingly used to educate outside of classrooms and within homes – often in the form of children using tablet computers under the purview of parents and caregivers. From the time of their first introduction to the market, the relatively low price of tablet computers contributed to their rapid adoption worldwide, with the research firm Gartner estimating 195 million devices sold in 2013 alone (Lunden, 2014). This has led the use of tablets for educational purposes in homes as well as in schools and device brands actively target parents for the sale of devices.

As was the case of the initial adoption of past technologies for educational purposes in schools and in homes, there is little formalized direction regarding the use of these devices – questions such as which tablet computers are most appropriate for which setting; which applications (software programs running on the tablet computers) best lead to preferred outcomes; or how to teach users to operate tablet computers - remain unanswered. Trial-and-error predominates and users base choices regarding devices and applications from their personal interests in a particular topic, cost (where free applications are especially attractive), word of mouth, or as a result of marketing efforts on- and off-line. Since users and particularly young users are often left to engage with the devices in a non-directed manner, we replicate this practice of non-directed user engagement with tablets in our study design as we investigate the effect of using these devices on cognition. Thus, our study involved not instructing participants on how to interact with the applications during tablet computer use.

Allowing users to engage with tablet computers and applications in a non-structured manner focuses the inquiry on how the app on the tablet computer directs the user. Luhmann (1992) in his seminal essay on communication proposes that we can assess understanding in communicative encounters by analyzing what information was requested and what information was expected in return. In tablet computer-user communications the app, in this case classified as educational, presents visual information directing the user to perform a touch or gestural response on the surface of the device. Typically, the user must determine which elements presented on the screen represent salient information (i.e., what are they being directed to do), and then the user must choose and execute gestural actions to satisfy the informational request. Sweller (1994) also considered the impact that learning interactivity has on cognitive load, where complex learning activities can either over or under load working memory and affect the development of schemas. Therefore, in app-directed communicative exchanges there are two contributing factors of interest: (i) app complexity, and (ii) the user’s cognitive ability.

Research questions

Considering the wide spread adoption of tablet computers in education and the dearth of research in this area, we pose the following research questions:

RQ1. How successful are educational applications on tablet computer at directing user’s attention towards intrinsic and germane content? This is achieved by comparing simple and complex applications to determine which of these two approaches to app design is more successful. Successful engagement is operationally defined as user’s (in this case children) (i) attending more to intrinsic and germane content than extraneous content and (ii) self-reporting that they enjoy the content.

RQ2. To what extent does a user’s cognitive ability affect engagement with educational applications? Assess whether a child’s cognitive ability affects how they successfully engage with educational content on tablet computers. Investigate whether or not children with more available short-term memory, working memory, and better attentional control are better able to direct their cognition towards intrinsic and germane content and away from extraneous content?

RQ3. Do application complexity and children’s cognitive ability interact?
**Method**

**Participants**

Participants were 30, English-speaking children (13 male) in Grade 2 with a mean age of 7 years and 3 months (range: 6 years, 8 months to 7 years, 9 months). Participants were from a single school in a large Canadian city. Ethics approval was obtained from the ethics committees of both the University of Toronto and the Toronto District School Board. A recruitment letter and parental consent form was sent to the homes of every child in Grade 2. All participants with parental consent participated in the study and provided verbal assent immediately prior to their participation. Only data from neurologically typical children are included in the reported results—neurological state was determined by asking teachers if the participant was formally assessed as neurologically atypical (e.g., assessed as having/being: a mild intellectual disability, Downs Syndrome, on the autism spectrum). The study took place in the first half of the 2013-2014 academic year.

**Procedure**

In individually-administered sessions conducted in a quiet room on school grounds, participants completed two sets of tasks. For the first set of tasks, participants used educational mathematics applications on two tablet computers (i.e., iPad & LeapFrog LeapPad 2) while gaze data was recorded with a desktop mounted 60Hz FaceLab 5 eye tracker (see Figure 1). Using the eye tracker entailed a four-point calibration procedure in which participants look at the four corners of the tablet computer screen while the eye tracker triangulates their gaze position for each location. The calibration procedure was conducted twice for each participant, once per tablet computer. After a successful calibration, the researcher demonstrated how to begin each educational application and participants used each application for a total of two-minutes. This duration was chosen so that the data would capture children’s initial interaction with an educational application (i.e., a learning phase). Following the use of each application, participants’ engagement with the application was assessed using self-report. Half of all participants used the iPad first and half used the LeapPad first. The order of the applications on each tablet computer was also cross-balanced. For the second set of tasks, participants completed four cognitive measures assessing short-term memory, working memory, and attention. All participants completed the cognitive tasks in the same order.

**Materials**

**Tablet computers**

The two tablet computers in the study represent a sample from the range of devices commercially available at the time of data collection, and also represent tablet computers that are most likely to engage children. In a previous
study assessing children’s interaction with tablet computers (i.e., iPad, LeapPad, Acer Iconia Tab, and VTech InnoTab), it was found that tablet computers designed for a general user audience (iPad & Iconia Tab) appear to be more engaging—are played longer, are used in a more goal-directed manner, and are judged as more enjoyable by both researcher observation and user’s self-reports—than tablets specifically designed for use by children (e.g., LeapPad & InnoTab; McEwen & Dubé, 2015), with the iPad and LeapPad garnering the most engagement in their respective categories.

Educational mathematics applications

One goal of the present study is to determine how successful tablet computers are at directing children’s attention towards intrinsic and germane educational content (cf., extraneous content). In this context, child-tablet interaction is mediated through application use. To assess the role applications have in user engagement, two applications on each of the tablet computers were chosen. The applications in the present study were chosen because (a) in a previous study they were found to be more engaging than applications on competing tablet computers (McEwen & Dubé, 2015) and (b) the applications represented diametrically opposed approaches on how to engage users, simplicity vs. complexity. Simple applications focused on one type of mathematics content, contained one type of learning mechanic, and had relatively plain visuals. Complex applications contained multiple types of mathematics content and learning mechanics and used relatively dynamic visuals (see Table 1).

Measures

Participants’ engagement with each application was measured using eye tracking metrics and a self-report measure. The selection of eye tracking metrics reported in the present study was chosen from a broader range of eye tracking data automatically produced by the software Eyeworks. The self-report measure was modeled off of work by Fisher, Dobbs-Oates, Doctoraff, and Arnold (2012), who assessed children’s engagement in paper and pencil mathematics tasks. This measure indexes engagement with the hypotheses that engaging applications are more enjoyable to use.

Eye tracking measures of engagement

*Fixation count.* Fixations are moments of relative stability in the eye during which encoding occurs (Poole & Ball, 2006). Fixation count is the total number of fixations in a given area of interest (AOI), with more fixations indicating that the areas is more noticeable or important to the user than other areas (Poole, Ball, & Phillips, 2005).

*Fixation duration.* Fixation duration is the average length of an individual fixation. Longer fixation durations within an AOI indicate difficulty in extracting information from that area (Just & Carpener, 1976).

*Gaze.* Gaze is the sum of all fixation durations within an area (Poole & Ball, 2006). Gaze can be used to visually compare how attention is divided between multiple AOIs with a stronger gaze indicating that more attention is directed to one area over another (Mello-Thoms, Nodine, & Kundel, 2002).

Self-report of engagement

Immediately following the use of each application, participants were asked how much they liked the application by presenting them with a line bounded by two choices.

```
I did not like it-----------------------------------I liked it a lot
```

To indicate their attitude towards the application, participants were instructed to place a mark on the line between the two choices. The distance from the left-most position on the line to the mark was used as an index of participants’ self-report of engagement.
Table 1. Educational mathematics applications

<table>
<thead>
<tr>
<th>Application</th>
<th>Math content</th>
<th>Complexity</th>
<th>Example screen image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion Math Zoom</td>
<td>Number line estimation</td>
<td>Simple</td>
<td><img src="image1" alt="Screen Image" /></td>
</tr>
<tr>
<td>Number Land HD</td>
<td>Counting, number identification and production</td>
<td>Complex</td>
<td><img src="image2" alt="Screen Image" /></td>
</tr>
<tr>
<td>Dice Ahoy</td>
<td>Probability</td>
<td>Simple</td>
<td><img src="image3" alt="Screen Image" /></td>
</tr>
<tr>
<td>T-Rex Rush</td>
<td>Counting, magnitude comparison, number and symbol identification</td>
<td>Complex</td>
<td><img src="image4" alt="Screen Image" /></td>
</tr>
</tbody>
</table>

*Note. EXT = Extraneous, INT = Intrinsic, GER = Germaine.*

**Cognitive tasks**

Two tasks were used to assess the capacity of participants’ short-term memory, one task was used to assess working memory, and one task was used to assess attention.

*Memory tasks.* The forward digit span, spatial span, and reverse digit span tasks were used as measures of verbal short-term memory, visual short-term memory, and working memory, respectively (Alloway & Alloway, 2010; Gathercole & Pickering, 2000; Wiseheart, Altmann, Park, & Lomardino, 2009). For the forward and reverse digit
span, the researcher read aloud a series of single-digit numbers at 1 second intervals. For the forward digit span task, participants were instructed to repeat the digits in the same order in which they were presented, which requires the numbers to be stored in verbal short-term memory. In the reverse digit span task, participants were instructed to repeat the digits in the reverse order, which requires the number to be manipulated in working memory. For the spatial span task, twelve shaded and identical circles were presented on a piece of paper. The researcher pointed to the circles sequentially and the participant was to touch the same circles in the same order, which requires the series to be stored in visual short-term memory. The series of numbers and circles were initially three items long and increased by one item after every correct repetition by the participant (Werheid et al., 2002). No feedback was given throughout the tasks. Responses were recorded as either right or wrong. The task ended after the participant incorrectly repeated two series with the same number of items. The longest correctly repeated series was recorded as a participant’s score.

Attention task. Manly et al.’s (2001) Opposite World task from the Test of Everyday Attention for Children (TEA-CH) was used as a measure of controlled attention. Participants were presented with a stimulus sheet containing a mixed array of the digits 1 and 2 (see Figure 2). There were two conditions, same world and opposite world. In the same world condition, participants were asked to read the digits aloud as quickly as possible from start to finish. In the opposite world condition, participants were asked to say the opposite for each digit (saying 1 for 2 and 2 for 1) as quickly as possible, which requires attention to be directed at the goal at hand (saying the opposite) by inhibiting the prepotent response (e.g., saying 1 for 1). In the task, participants only progressed to the next digit following a correct response, thus errors were incurred as a time penalty. Following a practice in each condition, four test pages were run in the same world condition and then four pages were run in the opposite world condition. The amount of time to complete each page was recorded using a stopwatch. The difference in total completion time between the opposite world and the same world task was taken as the dependent variable, with smaller differences indicating better controlled attention.

Figure 2. Example trial from controlled attention task

Results

RQ1. How successful are educational applications on tablet computer at directing user’s attention towards intrinsic and germane content?

To determine how successfully applications direct children’s attention towards intrinsic and germane content and away from extraneous content, two 2 (complexity: simple, complex) x 3 (content: intrinsic, germane, extraneous) ANOVAs were performed on the fixation count and fixation duration data. Intrinsic content included visuals that were required to complete the learning task (e.g., numbers, symbols, counting manipulatives). Germene content included visuals that scaffold the learning task but was not necessary to complete the task (e.g., progress markers, avatars/characters that offered aid on request, animations that reinforced the learning task). Extraneous content included visuals that did not aid the learning task (e.g., interactive and static background visuals, navigational
buttons). Examples of the content areas can be found in Table 1. Five participants’ eye tracking data are excluded in
the analyses due to poor calibration.

For fixation count, there was a main effect of complexity with more fixations for the simple applications than the
complex applications, $F(1, 50) = 13.699$, $MSE = 1071.233$, $p = .001$. There was also a main effect of content type
with more fixations in the intrinsic and germane content than in the extraneous content, $F(2, 50) = 19.289$, $MSE =
1071.233$, $p < .001$. These results suggest that the simple applications are more noticeable/important than the
complex applications and that the intrinsic and germane content are more noticeable/more important to the children
than the extraneous content (see Figure 3).

![Figure 3. Fixation count by complexity and content (error bars = se)](image)

For fixation duration, there was a non-significant trend for the average fixation to be longer for the simple
applications than for the complex applications, $F(1, 50) = 2.973$, $MSE = .1$, $p = .09$. No other main effects or
interactions are significant. However, the graphed data do provide some insight into the trend (see Figure 4), with
longer fixations in the extraneous and germane content for the complex applications.

![Figure 4. Fixation duration by complexity and content (error bars = se)](image)

**RQ2. To what extent does a user’s cognitive ability affect engagement with educational applications?**

To determine whether cognitive abilities affect how successfully children engage with tablet computers in a learning
context, scores on the memory and attention tasks were used to identify homogenous groups of participants who
shared similar cognitive profiles. To this end, a K-means cluster analysis (a hierarchical clustering algorithm based
on Euclidian distances) was performed on participants’ performance on the forward digit span, reverse digit span,
spatial span, and controlled attention tasks. The cluster solution grouped participants into two relatively
homogeneous groups of cases (see Table 2), with significant differences between the two groups on the digit backward and controlled attention tasks (i.e., executive functioning tasks). Thus, the analysis identified a group of children who possess lower executive functioning ability (i.e., Low EF) and a group of children who possess higher executive functioning ability (i.e., High EF), relative to each other.

<table>
<thead>
<tr>
<th>Task</th>
<th>Score</th>
<th>r(23)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit Forward (items)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low EF</td>
<td>5.17 (1.03)</td>
<td>-0.151</td>
<td>.88</td>
</tr>
<tr>
<td>High EF</td>
<td>5.23 (1.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial Span (items)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low EF</td>
<td>3.83 (.38 )</td>
<td>-0.05</td>
<td>.96</td>
</tr>
<tr>
<td>High EF</td>
<td>3.85 (.80 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Backward (items)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low EF</td>
<td>2.58 (.51)</td>
<td>-3.592</td>
<td>.002*</td>
</tr>
<tr>
<td>High EF</td>
<td>3.54 (.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controlled Attention (seconds)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low EF</td>
<td>8.76 (5.97)</td>
<td>3.702</td>
<td>.007*</td>
</tr>
<tr>
<td>High EF</td>
<td>3.77 (.82)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *Scores on the memory tasks represent the total number of items recalled. Scores on the attention task represents the difference in completion time between the same world and opposite world tasks, with lower scores indicating better controlled attention.

To determine how successfully children direct their attention towards intrinsic and germane content and away from extraneous content, two 2 (cognitive ability: low EF, high EF) x 3 (content: intrinsic, germane, extraneous) ANOVAs were performed on the fixation count and fixation duration data. There was a main effect of content type with more fixations in the intrinsic and germane content than in the extraneous content, $F(2, 46) = 19.224, MSE = 2267.251, p < .001$. There was no main effect or interaction with cognitive ability, $F$s < 1.0. The main effect suggests that the intrinsic and germane content is more noticeable/important to the children than the extraneous content whereas an inspection of the graphed data suggests that the high EF children find the intrinsic content more important than the germane content, the same is not true for the low EF children (see Figure 5).

![Figure 5. Fixation count by cognitive ability and content (error bars = se)](image)

For fixation duration, contrasts revealed that there was a non-significant Content X Cognitive Ability interaction, $F(1, 23) = 3.615, MSE = .011, p = .07$. The graphed data suggests that low EF children are having more difficulty extracting information than the high EF children, low EF children exert more effort to extract information from the extraneous content, and high EF children exert less effort to extract information from the extraneous content (see Figure 6).
RQ3. Do application complexity and children’s cognitive ability interact?

To determine whether the interaction between application complexity and cognitive ability differentially affect whether attention is directed towards intrinsic and germane content and away from extraneous content, two 3 (content: intrinsic, germane, extraneous) x 2 (complexity: simple, complex) x 2 (cognitive ability: low EF, high EF) ANOVAs were performed on the fixation count and fixation duration data. For fixation count, the Content X Complexity X Cognitive Ability interaction was significant, $F(2, 46) = 5.593$, $MSE = 927.832$, $p = .007$. The graph data suggests that low EF children rely more on the germane content for aid in the simple applications but rely less on the germane content in the complex applications (see Figure 7). In contrast, the high EF children do the opposite—they rely more on the germane content for aid in the complex applications and rely less on the germane content for aid in the simple applications (see Figure 7). For the fixation duration data, the three-way interaction did not approach significance, $F < 1.0$.

To further assess the three-way interaction using the eye tracking data and to present a more concrete representation of the eye tracking data, heat maps of gaze data were generated depicting the sum of all fixation durations within an area. This data was generated from a subset of participants ($n = 9$) because the calibration procedure often required the tablet computer to be moved and this prevents the amalgamation of gaze data due to the tablet computer being physically located in different areas of the eye tracker’s scene camera. The physical location of the tablet computer does not affect the other reported analyses, in which the AOIs are individually created for each participant and then used to amalgamate the data. This data can be used to visually compare how attention is divided, with a stronger gaze (indicated in a heat map as red) suggesting that more attention is directed to one area over another (Mello-Thoms et al., 2002; see Table 3). The gaze data heat maps suggest that low EF children’s attention is more divided for simple
applications and more focused for complex applications. In contrast, high EF children’s attention is more focused for simple applications and more divided for complex application. This data supports the previous three-way interaction in the fixation count data in that high EF children take advantage of the germane content in complex applications whereas low EF children seem to take advantage of the germane content for simple applications.

Table 3. Heat maps of gaze data for low and high EF clusters by application complexity

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Low EF (n = 5)</th>
<th>High EF (n = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 8. Enjoyment by cognitive ability and content (error bars = se)

The previous analyses of the eye tracking data inform how application complexity and cognitive ability affect how children engaged with the educational content on a tablet computer but do not necessarily provide information on how the engagement is experienced by children. To this end, a 2 (complexity: simple, complex) x 2 (cognitive ability: low EF, high EF) ANOVA was performed on children’s self-report of engagement. A significant Complexity X Cognitive Ability interaction indicates that low EF children enjoy the simple applications more than the complex applications.
applications whereas high EF children have no preference, $F(1, 22) = 5.994$, $MSE = 717.133$, $p = .023$ (see Figure 8). This suggests that low EF children enjoy applications that are in line with their cognitive ability.

**Discussion**

The tablet computer applications were able to direct children’s attention to the intrinsic and germane content. From a Luhmann-communications perspective, the design of the content in the applications promoted a successful user-tablet computer communicative encounter because the more educationally cogent information was more visually attractive leading to more successful child-tablet computer interaction, regardless of the children’s cognitive abilities.

From a cognitive psychology perspective an interpretation of this result is that the tablet applications all generate high information interactivity—that is, information content cannot be learned in isolation but must be understood by their relations to each other on screen (Sweller, 1994 p. 304-306). For Sweller (1994) in learning scenarios where there is high interactivity—as is the case in the multimedia content found on tablet computers—extraneous cognitive load can interfere with learning. The finding that the extraneous content was less “noticeable” by children may indicate that children were only manipulating and developing schemas for the intrinsic and germane content and did not have working memory available to process extraneous content. Thus, it is both a function of intentionally sound application design and an indication of high interactivity in the educational application learning context.

Another result of interest is that the children assessed as high EF found the intrinsic content from the educational applications more important than the germane content, while low EF children found the germane content more important than the intrinsic content. This could indicate that high EF children were utilizing pre-existing schemas in the processing of intrinsic content (hence a lower germane load), while low EF children needed to create and/or automate schemas to process the intrinsic content encountered in the applications (hence a higher germane load) (Paas et al., 2004).

Two results that appear to be consistent with existing theory from both communications studies and cognitive psychology are that, (a) low EF children appeared to experience more difficulty extracting information from the applications than high EF children; and (b) low EF children seem to take advantage of the germane content for simple applications but high EF children take advantage of the germane content for complex applications. In both (a) and (b) the analysis could be the same, and both communication theorists and cognitive psychologists could agree that children’s success in managing and extracting information content is positively related to their cognitive abilities. From a communications theory perspective, more successful user-device communication (messages from the sender is received and understood by the receiver) is indicative of higher cognitive ability. From a cognitive psychology perspective, increased executive functioning is co-related with the child’s utilization of existing schemas (i.e., attending to intrinsic content) or the creation and utilization of new schemas (i.e., attending to germane content), when required by a high level of information interactivity (i.e., complex applications).

**Conclusion**

The primary goal of this study was to explore the relationship between a user’s engagement with tablet computers and user’s cognitive load, using theories and methodologies from communications studies and cognitive psychology. This goal was attained and the study demonstrates the value of employing an interdisciplinary approach to the study of new media.

There is evidence of the validity of Botha’s (1992) claim that the cognitive structure of the individual is intimately linked to the forms or systems of communication used. The research showed that tablet computers and their applications offer a learning experience that appears to be inherently highly interactive and thereby introducing challenges to the cognitive load of children as users. More research is needed to determine whether this finding is generalizable to adults, and a broader range of applications and tablet computers could be investigated to see whether or not subject matter affects the interactivity of information elements in the applications. However, this offers a start in the development of a broader theory on cognitive load and touch devices. An extension of Botha’s claim arising from this study is that cognitive abilities/structures may be linked to the forms of communications used, but so is the
reverse: the forms of communications are also defined and linked to our cognitive abilities to interact effectively. This offers an example of the co-constitutive nature of media and use.

The use of an eye tracker in data collection was instrumental to the success of this study—many of the results would not have been derived without its use. We hope to encourage other researchers to include this method as part of the data particularly as it allowed for the assessment of vision.

Designers of educational applications should consider these findings in the development of new applications for children. In particular, they could take into account the deleterious cognitive effect of the complex applications on children with lower EF, applications that are typically marketed as being able to engage children who have difficulty learning in “traditional” scenarios. This means that designers could conduct better app testing and invite test participants not just based on age or grade levels, but also include participants with a range of executive functioning skills to offer a better outcome for all users.

Acknowledgements

Funding for this research was provided by the following: Canadian Foundation for Innovation, Social Science and Humanities Research Council, and the University of Toronto’s Faculty of Information. Thank you to Vanessa Rowlin and Akash Venkat for assistance with the collection and coding of the data.

References


Do Learner Characteristics Moderate the Seductive-Details-Effect? A Cognitive-Load-Study Using Eye-Tracking

Babette Park*, Andreas Korbach and Roland Brünken

Department of Education, Saarland University, Saarbrücken, Germany // b.park@mx.uni-saarland.de // a.korbach@mx.uni-saarland.de // r.bruenken@mx.uni-saarland.de

*Corresponding author

ABSTRACT

The present study examines whether the seductive-details effect is moderated by spatial ability and prior knowledge, which are two of the most relevant learner characteristics in multimedia learning. It is assumed that the seductive-details effect with an increase in extraneous cognitive load and a decrease in perceptual processing and learning success is only present for learners with low spatial ability and low prior knowledge. To this end, the present study uses an Aptitude-Treatment-Interaction Design with separate analyses for spatial ability and prior knowledge as aptitude-variables and seductive details (with vs. without) as treatment-variable. Participants (N = 50) were asked to learn about biology with a multimedia instruction that manipulated seductive details. The results show that learners perceptual processing, measured by eye-tracking, and learning performance was significantly lower when learning with seductive details. In addition, spatial ability and especially prior knowledge were confirmed to play the expected moderating role.

Keywords

Seductive details, Multimedia learning, Spatial ability, Prior knowledge, Eye-tracking

Introduction

The design of computer-based multimedia learning instructions has many options concerning the integration of additional, non-redundant and interesting but irrelevant learning material in form of pictures, text, animated sequences, videos or audio commentaries. These options are more than playing with colors and/or shapes of the relevant learning material as recommended by emotional design principles that can evoke learning-conducive affective processing in multimedia learning (Park, Knörzer, Plass, & Brünken, 2015; Park, Plass, & Brünken, 2014; Plass, Heidig, Hayward, Homer, & Um, 2014). The additional, non-redundant and interesting but irrelevant information is also used to make the learning material more interesting and attractive to learners (Park, Flowerday, & Brünken, 2015). However, in fact such additional information can also decrease the learning performance.

Until now, research on this negative effect of seductive details has focused on seductive text passages or seductive illustrations in text comprehension studies. Several studies have shown a detrimental effect of seductive details (Garner, Gillingham, & White, 1989; Harp & Mayer, 1998; Lehman, Schraw, McCrudden, & Hartley, 2007; McCrudden & Corkill, 2010), whereas others have shown non-significant results (Garner & Gillingham, 1991; Hidi & Baird, 1988). All of these studies showing a detrimental effect were using scientific texts that explain for example detailed differences between insects or the lightning process step by step. In contrast, the studies that could not show the detrimental effect of seductive details were using non-scientific text. This already is a hint to the idea that seductive details can only interfere with learning within a high-loading learning process that requires managing the available cognitive resources.

In a study by Park, Moreno, Seufert and Brünken (2011), it was shown that controversial results in seductive-details research can be explained by an effect on cognitive load. The findings showed that students' learning performance was significantly higher when seductive details were presented under the low load condition (narration) as compared to all other conditions. Concerning particular learner characteristics a similar effect should appear: If the degree of cognitive load is responsible for the strength of the seductive-details effect and the individual degree of cognitive load is affected by learner characteristics, there should be learner characteristics that moderate the seductive-details effect. To this end, the goal of the present study was to test this hypothesis for two learner characteristics that are supposed to affect the individual cognitive load while learning with multimedia learning instructions: prior knowledge and spatial ability.
Theoretical framework and predictions

According to the Cognitive Load Theory (CLT) (Plass, Moreno, & Brünken, 2010; Sweller, Ayres, & Kalyuga, 2011) the total cognitive capacity is limited and the amount of total cognitive load is determined by three components. First, intrinsic cognitive load depends on the element interactivity. The larger the number of elements that must be processed in working memory and the more complex their relation to each other is, the higher the intrinsic load. Second, extraneous cognitive load is directly caused by the format and concept of the information presentation. A proper instructional design fosters information processing and saves cognitive resources by minimizing extraneous cognitive load. Third, germane cognitive load is the load dedicated to relevant information processing. The higher the engagement in learning and schema acquisition is, the higher the germane load. Seductive details consist of additional interesting but irrelevant information, are part of the instructional design and can therefore be allocated to the extraneous load factor. So adding seductive details to a learning content causes additional extraneous load and may overstrain the learners’ cognitive capacity. Within an extraneously low loading instructional design (e.g., audio-visual learning material), participants may have enough capacity to process the relevant learning content and the seductive details whereas in an extraneously high loading instructional design (visual-only learning material) the processing of additional information decreases learning performance (Park et al., 2011). However, there is still the question about how exactly seductive details affect information processing and learning within an extraneously high loading learning instruction as often found in learning environments of schools and universities (e.g., text books, handbooks, hypertexts including pictorial information).

Theoretical explanations for the seductive-details effect

Harp and Mayer (1998) provide three explanations for the seductive-details effect, the diversion, disruption or distraction of the relevant learning process. First, the diversion hypothesis assumes that seductive details activate inappropriate prior knowledge and cause inappropriate schemata that are organized by integrating the new information with the activated prior knowledge. Some studies tested the effect of schema interference by manipulating the presentation order of seductive details in the way that seductive details were presented at the beginning, interspersed, or at the end of the learning material (Harp & Mayer, 1998). As seductive details did only affect learning in a negative way, when presented before or within the learning session and not when presented after the learning session, the results support the diversion hypothesis and the assumption of schema interference. In case of the presentation of seductive details before or early in the learning session, students try to integrate this information or use the seductive-details information like an anchor for the whole learning session leading to the wrong focus of information perception and processing during the learning session. Second, the disruption hypothesis assumes a coherence disruption of the relevant information processing by seductive details. A study by Lehman et al. (2007) gives support for this assumption showing that seductive details reduce reading time of relevant sentences in scientific text and decrease the recall of main ideas. Third, the distraction hypothesis assumes an attention distraction of the relevant information processing by seductive details. Sanchez and Wiley (2006) investigated the influence of working memory capacity, as a learner characteristic, on the effect of seductive illustrations and animations. Learners with low working memory capacity were significantly more disturbed by seductive details (illustrations), than those with a higher memory capacity and drew their attention more often and for longer time intervals to seductive details, as registered by eye-tracking. In sum, cognitive capacity seems to be of great importance to explain the seductive-details effect and so should learner characteristics be very important for the explanation of the seductive-details effect because they influence the use of cognitive resources during the learning process, too.

Prior knowledge and spatial ability

A very important learner characteristic assumed to affect the effect of instructional multimedia design principles on learning success is the learners’ prior knowledge (Kalyuga, Chandler & Sweller, 1998). Research on the expertise reversal effect shows that the application of several supportive instructional design principles do only increase the learning success for learners with low prior knowledge, whereas learning success for learners with high prior knowledge is unaffected or even decreased (Kalyuga, Ayres, Chandler & Sweller, 2003). Thus, the assumption suggests itself that learners with high prior knowledge are affected in another way than learners with low prior knowledge, when learning with seductive details. According to CLT, element interactivity can be decreased by domain-specific knowledge and the intrinsic cognitive load for learners with high prior knowledge should be lower
than for learners with low prior knowledge. As mentioned above, seductive details affect the learning success especially under high load conditions (Park et al., 2011), so the seductive-details effect should affect learners with low prior knowledge in a significantly stronger way than learners with high prior knowledge. A study by Magner, Schwonke, Aleven, Popescu, and Renkl (2014) gives clear support for this assumption.

Another important learner characteristic is the learners’ spatial ability. Especially while learning with a multimedia instruction that uses visual-figural and three-dimensional-spatial information, the learners’ spatial ability is of great importance for the construction of three-dimensional mental representations out of two-dimensional visual figural information (Münzer, Seufert, & Brünken, 2009). For instance, in a study by Mayer and Sims (1994) it was shown that the contiguity effect is strong for high spatial ability participants, thus only the high spatial ability learners performed better under concurrent presentation of narration and animation. This finding supports the assumption that spatial ability fosters the construction of a mental model and that building such a visual representation is much more demanding for low spatial ability learners. Thus, high spatial ability learners should be less affected by seductive details because handling a three-dimensional mental animation should be less demanding and should cause less intrinsic cognitive load. As the focus of attention is very important in order to construct such three-dimensional mental models when learning with multimedia instruction, we also assume spatial ability to affect learners’ focus of attention. Thus, a detrimental effect of seductive details and a moderating influence of prior knowledge and spatial ability is not only expected for the learning performance but also for the focus of attention, indicated by eye-movement data. With respect to the study by Magner et al. (2014), the present study realizes a closer look on the moderating influence on learning success by a detailed analysis of learners’ information processing.

Eye-tracking the seductive-details effect

To get a closer look on the influence of seductive details on the learners’ information processing, eye-tracking methodology can be a very useful tool. As it’s indicated by several studies there is evidence for a close relation between eye-movement measures like total fixation time and cognitive activity that supposes e.g., long fixation time as an indicator for high cognitive activity (Just & Carpenter, 1976). Moreover, the total fixation time on the relevant picture in multimedia learning is hypothesized to cause cognitive processing and to serve as a measure of cognitive performance (Rayner, Li, Williams, Cave & Well, 2007). Eye-tracking provides information about the perceptual processing while learning and in combination with measures of learning success it provides information about cognitive processing (Mayer, 2010). If particular learner characteristics have a moderating influence on the seductive-details effect, these learner characteristics should also affect the learners’ eye movement.

As mentioned above, seductive details consist of additional but irrelevant information. Concerning eye-tracking research, Canham and Hegarty (2010), for instance, found an effect on the ability to focus task-relevant information dependent on the participants’ domain knowledge, in the way that higher knowledge enhanced information selection. The learners’ prior knowledge should not only reduce intrinsic load but also enhance the information selection. Therefore, perceptual and cognitive processing should be affected by the learners’ prior knowledge and the learners’ prior knowledge should moderate the seductive-details effect, indicated by measures of perceptual processing.

For a better understanding of the seductive-details effect it could be useful to analyze when the learners look at relevant information for the first time. An appropriate eye-tracking measure for assessing the learners starting point of the information processing is the time to first fixation. This measure of perceptual processing provides information about the order of the learners’ fixations and about the time the learners start processing particular relevant information (Hyönä, 2010). One possible detrimental effect of seductive details is the learners’ attention distraction of cognitive processing the relevant information (Garner et al., 1989). If the presentation format of relevant and seductive-details material is held constant, the learners’ first fixations provide information about the influence of seductive details on their primary attention focus during the first seconds of the learning process.

In addition, eye-tracking data about the fixations on relevant pictorial information are assumed to provide information about the cognitive activity caused by the construction and handling of a visual mental representation. A process close to the handling of mental models and also related to spatial ability is the integration of verbal and figural information. Especially the transitions between semantically related text and pictorial information are assumed to indicate the cognitive engagement during the integration process of verbal and figural information (Holsanova, Holmberg & Holmqvist, 2009; Schmidt-Weigand, Kohnert & Glowalla, 2010). Thereafter, the influence
of the learners’ spatial ability on cognitive processing while learning with multimedia instruction should be indicated by several measures of perceptual processing in the form of a moderating effect, when analyzing seductive-details material.

**Goal of the present study**

The present study examines whether the seductive-details effect is moderated by memory associated learner characteristics such as spatial ability and by expertise related factors such as prior knowledge. It is assumed that processing seductive details will result in a decrease of the learning success. Furthermore it is assumed that learning with seductive details also affects the learners’ cognitive load. The learners who are working with the seductive-details version should experience significantly higher cognitive load in comparison to those working with the no-seductive-details version. If the seductive-details effect is moderated by spatial ability and prior knowledge, it is assumed that the seductive-details effect with an increase in cognitive load and a decrease in learning success is only present for learners with low spatial ability and low prior knowledge.

The present study further examines whether seductive details affect perceptual processing and whether this seductive-details effect is moderated by the learners’ prior knowledge and spatial ability. The seductive-details effect should be based on a decrease of the total fixation time and the total number of fixations on the relevant pictures, as well as the transitions between relevant textual and pictorial information. Furthermore it is assumed that seductive details also affect the learners’ first attention focus during perceptual processing. The learners who are working with the seductive-details version should fixate the relevant information later than those working with the no-seductive-details version. If the seductive-details effect is moderated by prior knowledge and spatial ability, it is assumed that the seductive-details effect with a decrease in factors of perceptual processing is only present for learners with low prior knowledge and low spatial ability.

**Method and data sources**

**Participants and design**

In order to assess the seductive-details effect on perceptual processing and learning success and the moderating role of learner characteristics an aptitude-treatment-interaction design was used, spatial ability and prior knowledge served as aptitude-variables and seductive details (with vs. without) as treatment-variable. The moderating effects of spatial ability and prior knowledge were assessed in separate analyses, as there was no significant correlation found between both aptitude-variables ($r = .263$, ns). Participants were 50 psychology students from a German University (79.6% female, average age = 22.1 years, $SD = 3.0$). They were randomly assigned to one of the two experimental groups.

**Materials**

Both groups worked with a self-directed multimedia-learning program that consists of 11 screens presenting an instruction about the structure and function of the ATP Synthase, a cellular molecule responsible for synthesis of ATP. The relevant information was presented as a combination of static pictures (see Figure 1, top left on each screen) and corresponding textual explanations (see Figure 1, below left on each screen). The objective of the learning task was to achieve a deep understanding of the molecule structure and the single steps in the process of ATP synthesis by integrating the verbal and pictorial representations (text = 440 words all pages together). All participants were introduced to the learning objective at the beginning of the learning task.

The experimental group worked with the seductive-details version, presenting additional and highly interesting but irrelevant information on 4 of the 11 screens in form of illustrations and related text on the right side of the screen (see Figure 1). In contrast to the relevant information, seductive details provided information about the usefulness of ATP that was not part of the learning objective. According to former studies seductive details were chosen by the following aspects: interestingness, irrelevance, concreteness, conciseness, emotionality and reference to the relevant topic (Garner et al., 1989; Park et al., 2011; Park, Flowerday, & Brünken, 2015; Sanchez & Wiley, 2006).
Figure 1. Example screens of the learning environment for the experimental group (seductive-details version, see left example screens) and the control group (without seductive details, see right example screens); original version in German, translated by the authors.

Measures

Working memory capacity measured by the numerical-memory-updating subtest of Oberauer, Süß, Schulze, Wilhelm, and Wittmann (2000), time-on-task, registered automatically by the computer, and participants learning motivation, measured by a revised short version of the 100-item Inventory of School Motivation (ISM; McInerney & Sinclair, 1991) Cronbach’s α = 0.86, served as control measures.

Prior knowledge was measured by a questionnaire that included five multiple-choice and eight open-ended questions, Cronbach’s α = 0.86. Spatial ability was measured by a standardized paper-folding and card-rotation test (Ekstrom, French, Harman, & Dermen, 1976).

Learning success was assessed by a learning-performance test including 12 items. The difficulty of each item lies between p = .20 and .80. The differentiation between two levels of required cognitive processing was considered by using the two subscales retention and comprehension. The subscale retention included 5 items, 3 in multiple-choice format and 2 in open response, showing a Cronbach’s α of 0.71. The subscale comprehension included 7 items, 3 in multiple-choice format and 4 in open response, showing a Cronbach’s α = 0.85.

The participants’ eye movements were recorded with a remote eye-tracking system (Tobii-TX300). Areas of interest (AOI) were defined only for relevant text and relevant pictures on each screen of the learning instruction (see Figure 1, relevant text and pictures on the left side of the screens) and not for the seductive-details material on the right side of some of the screens. The analyses focused on the total fixation time on the relevant picture, the total number of
fixations on the relevant picture, the time to the first fixation on the text and picture AOIs, and the learners’
transitions from text to picture AOIs.

In addition, total cognitive load was measured by subjective ratings (Paas, 1992). Participants were asked to rate
their cognitive load in the middle of the lesson (after screen 4 of 11) and immediately after the lesson on a seven-
point Likert-scale. The mean of both ratings was used as subjective ratings of cognitive load in the following
analyses.

Procedure

Participants started with the working memory capacity test, followed by the tests of spatial ability and prior
knowledge. The experimental group worked with the seductive-details version and the control group with the no-
seductive-details version of the multimedia instruction. Eye movements were recorded while learning. Subjective
ratings on cognitive load were collected once after learning screen four and once at the end of the learning
instruction. Finally, participants completed the learning performance test.

Results and conclusions

The two groups did not differ significantly concerning prior knowledge, $F(1,48) = 1.07, ns$, spatial ability, working
memory capacity, time-on-task, or learning motivation, $Fs < 1$, (see Table 1). In addition, the first screen of the
learning program shows only text, that is the same for all participants and was used to control the eye-movement
variables. There were no significant differences between the groups concerning the number of fixations or total
fixation duration, $Fs < 1$. In addition, independent samples $t$-tests were conducted for learning success, eye
movement and cognitive load with the between subject factor seductive details (with vs. without).

Table 1. Means and standard deviations for all control variables

<table>
<thead>
<tr>
<th></th>
<th>No seductive details ($n = 25$)</th>
<th>Seductive details ($n = 25$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge (max. = 13)</td>
<td>3.64(3.19)</td>
<td>4.62(3.5)</td>
</tr>
<tr>
<td>Spatial ability (%)</td>
<td>75.80(12.84)</td>
<td>72.80(17.94)</td>
</tr>
<tr>
<td>Working memory capacity (max. = 6)</td>
<td>3.88(1.4)</td>
<td>3.92(1.1)</td>
</tr>
<tr>
<td>Time-on-task (min.)</td>
<td>7.15(3.53)</td>
<td>7.47(3.25)</td>
</tr>
<tr>
<td>Motivation (max. = 5)</td>
<td>3.67(.25)</td>
<td>3.6(3.3)</td>
</tr>
<tr>
<td>Number of fixations for slide number 1 (N)</td>
<td>187.04(97.2)</td>
<td>162.17(109.3)</td>
</tr>
<tr>
<td>Total fixation duration for slide number 1 (sec.)</td>
<td>56.8(36.04)</td>
<td>50.01(37.84)</td>
</tr>
</tbody>
</table>

The first $t$-test shows that comprehension performance was significantly lower in the seductive-details group, $t(48) = 2.45, p = .009, d = .71$. No significant difference was found for retention performance, $t(48) = .278, ns$ (see Table 2). This result confirms the seductive-details effect on a higher level of required cognitive processing measured by the comprehension-subscale.

When analyzing the eye-movement behavior of the experimental group in contrast to the control group on the slides
where seductive details appear for the experimental group, different effects were found. First of all, learners showed
significantly shorter total fixation times on the relevant pictorial information in the seductive-details group in
contrast to learners who learned without seductive details, $t(43) = 1.806, p = .039, d = .55$. Second, learners of the
seductive-details group executed significantly fewer fixations on the relevant pictorial information, $t(43) = 2.234, p = .015, d = .68$. Third, the seductive-details group fixated the relevant pictorial information significantly later than the group without seductive details, $t(42) = -2.412, p = .010, d = .74$. Finally, learners of the seductive-details group demonstrated significantly fewer transitions between the relevant textual and pictorial information in contrast to learners who learned without seductive details, $t(43) = 3.253, p = .001, d = .99$ (see Table 2).

Moreover, learners of the seductive-details group rated their cognitive load to be significantly lower in contrast to
learners of the no-seductive-details group, $t(48) = 1.83, p = .036, d = .53$ (see Table 2).
Table 2. Means and standard deviations for all variables

<table>
<thead>
<tr>
<th></th>
<th>No seductive details (n = 25)</th>
<th>Seductive details (n = 25)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M(\text{SD}))</td>
<td>(M(\text{SD}))</td>
</tr>
<tr>
<td>Comprehension (max. = 16.5)</td>
<td>10.6(3.0)</td>
<td>7.9(4.7)</td>
</tr>
<tr>
<td>Retention (max. = 12)</td>
<td>7.54(2.2)</td>
<td>7.32(3.3)</td>
</tr>
<tr>
<td>Cognitive load (max. = 7)</td>
<td>5.1(.93)</td>
<td>4.4(1.6)</td>
</tr>
<tr>
<td>Total fixation duration on the relevant pictures =</td>
<td>50.9(33.9)</td>
<td>34.8(25.3)</td>
</tr>
<tr>
<td>Picture AOIs (sec.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total fixation count on the relevant pictures =</td>
<td>204.8(158.25)</td>
<td>121.6(81.3)</td>
</tr>
<tr>
<td>Picture AOIs (N)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitions from relevant text to relevant picture =</td>
<td>23.7(16.3)</td>
<td>11.9(7.7)</td>
</tr>
<tr>
<td>Transitions between text and picture AOIs (N)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to first fixation (sec.)</td>
<td>0.26(0.31)</td>
<td>2.06(3.49)</td>
</tr>
</tbody>
</table>

In order to investigate the moderating role of spatial ability or prior knowledge, respectively, on the seductive-details effect, separate moderation analyses were conducted. Spatial ability or prior knowledge served as moderator and learning success, eye-movement data or subjective ratings of cognitive load were used as dependent variables. The conducted moderation analysis is a regression-based approach for conditional process modeling by Hayes (2013).

The first analysis assesses the moderating influence of spatial ability on the seductive-details effect concerning learning success. A moderation analysis was conducted for comprehension with spatial ability as moderator. The regression model was significant, \(F(3,45) = 2.8, R^2 = .16, p = .050\). In accordance with the result of the t-tests, the regression analysis shows a main effect for seductive details, \(t(45) = -2.08, \beta = -1.11, p = .043\), but no main effect for spatial ability, \(t(45) = 1.37, \beta = 5.09, ns\) and no interaction effect, \(t(45) = .72, \beta = 2.65, ns\). The regression coefficients show (marginal) significant conditional effects for the 10th, 25th, and 50th (but not for the 75th and 90th) percentiles of spatial ability, with \(\beta = -1.66, p = .084; \beta = -1.41, p = .045;\) and \(\beta = -1.12, p = .043\), indicating that learners with low levels of spatial ability are more affected by seductive details (see Figure 2).

![Figure 2. Comprehension moderated by spatial ability](image)

In purpose of assessing the moderating influence of prior knowledge on the seductive-details effect concerning learning success, a moderation analysis was conducted for comprehension. The regression model was significant, \(F(3,46) = 3.8, R^2 = .20, p = .016\). In accordance with the result of the t-test, the regression analysis shows a main effect for seductive details, \(t(46) = -2.8, \beta = -1.53, p = .007\). There was also a main effect for prior knowledge, \(t(46) = 2.08, \beta = .34, p = .042\), but no interaction effect, \(t(46) = .66, \beta = .11, ns\). The regression coefficients show (marginal) significant conditional effects for the 10th, 25th, 50th, and 75th (but not for the 90th) percentiles of prior knowledge,
with $\beta = -1.98$, $p = .027$; $\beta = -1.87$, $p = .016$; $\beta = -1.65$, $p = .006$; $\beta = -1.27$, $p = .065$, indicating that learners with low prior knowledge are more affected by seductive details (see Figure 3).

To assess the moderating influence of spatial ability on the seductive-details effect concerning perceptual processing, an additional moderation analysis was conducted for one measure of eye movement with spatial ability as moderator. Because the transitions between text and picture AOIs are the most relevant indicator for perceptual processing as well as for integration processes and therefore most interesting when analyzing the aptitude-treatment-interaction between spatial ability and seductive details, the following results focus on this indicator.

The regression model was significant, $F(3,40) = 3.2$, $R^2 = .19$, $p = .034$. In accordance with the result of the $t$-test, the regression analysis shows a main effect for seductive details, $t(40) = -3.07$, $\beta = -5.99$, $p = .003$, but no main effect for spatial ability $t(40) = -5.0$, $\beta = -7.06$, ns, and no interaction effect $t(40) = .51$, $\beta = 7.36$, ns. The regression coefficients show (marginal) significant conditional effects for the 10th, 25th, 50th, and 75th (but not for the 90th) percentiles of spatial ability, with $\beta = -7.46$, $p = .042$; $\beta = -6.77$, $p = .011$; $\beta = -5.94$, $p = .004$; $\beta = -4.92$, $p = .081$, indicating that learners with low levels of spatial ability are more affected by seductive details (see Figure 4). Learners of the no-seductive-details group show fewer transitions the higher their spatial ability even though they show the same high learning performance. However, when learning with seductive details, this advantage due to high spatial ability seems not to be present: Learners show the same integration activity by using transitions from relevant text to relevant pictures with low or high spatial ability, but their learning performance increases with increasing spatial ability (compare Figure 2 and 4). Thus, seductive details reduce the integration activity and processes and hinder learners with low spatial ability to reach a high learning performance. However, seductive details do not hinder learners with high spatial ability to reach a high learning performance, while showing a comparable low investment of integration processing to learners with low spatial ability.

In order to assess the moderating influence of prior knowledge on the seductive-details effect concerning perceptual processing, a separate moderation analysis was conducted for one measure of eye movement with prior knowledge as moderator. Transitions between text and picture AOIs are again the most relevant indicator for perceptual processing as well as integration processes and therefore also chosen for the following analysis.

The regression model focusing on the transitions between text and picture AOIs was significant, $F(3,41) = 4.7$, $R^2 = .25$, $\beta = -5.62$, $p = .006$. In accordance with the result of the $t$-test, the regression analysis shows a main effect for seductive details, $t(41) = -3.01$, $\beta = -5.62$, $p = .004$, but no main effect for prior knowledge, $t(41) = -1.66$, $\beta = -.94$, $ns$, and no interaction effect, $t(41) = .84$, $\beta = .48$, $ns$. The regression coefficients show (marginal) significant conditional effects for the 10th, 25th, 50th, and 75th (but not for the 90th) percentiles of prior knowledge, with $\beta =$
7.66, \( p = .014 \); \( \beta = 7.18, \ p = .008 \); \( \beta = 6.22, \ p = .003 \); and \( \beta = 4.55, \ p = .055 \) indicating that learners with low levels of prior knowledge are more affected by seductive details (see Figure 5).

**Figure 4.** Transitions between text and picture AOIs moderated by spatial ability

**Figure 5.** Transitions between text and picture AOIs moderated by prior knowledge

In purpose of assessing the moderating influence of spatial ability or prior knowledge, respectively, on the seductive-details effect concerning cognitive load, two separate moderation analysis were conducted. Spatial ability served as moderator in the first analysis. The regression model was not significant, \( F(3,45) = .82, R^2 = .05, \ ns \). Prior knowledge was chosen as moderator in the second analysis. The regression model was significant, \( F(3,46) = 5.3, R^2 = .26, \ p = .003 \). Results show an effect for prior knowledge, \( t(46) = -3.28, \ \beta = -.17, \ p = .002 \), with higher cognitive load ratings for learners with low prior knowledge, but no effect for seductive details, \( t(46) = -1.5, \ \beta = -.26, \ ns \), no interaction effect, \( t(46) = -.79, \ \beta = -.04, \ ns \), and no conditional effects of the moderator.
**Discussion**

**Seductive-details effect due to disrupting processes**

In sum, the results confirm our hypothesis and show a detrimental effect of seductive details not only on learning success but also on perceptual information processing. The eye-tracking data show that seductive details influence mainly the processing of relevant pictorial information and cause a less deep information processing. The seductive-details group fixated the relevant pictorial information for a shorter time interval, less frequently in contrast to the no-seductive-details group and executed less integrative transitions from text to picture AOIs. These findings are very interesting because time-on-task was not strictly controlled in this study, as learners only had to be on the screens for at least the minimum time (empirically tested minimum reading time of the screens) and could be on the screen for a given maximum time (approximately 2 minutes). Learners in the seductive-details group had the possibility to process the relevant information with the same extent as the group without seductive details. However, seductive details affected their focus of attention and especially their allocation of processing time. Thus, the results give support for the disruption hypothesis (Lehman et al., 2007) with a harmful effect on deeper processing and a reduction of processing time for the relevant information. Further research and a more detailed analysis of the eye-tracking data is needed to assess the disruption hypothesis and to answer the question if seductive details indeed interrupt the processing of the relevant information and disrupt coherence formation. As the integration of textual and pictorial information was crucial for learning success, transitions between relevant and seductive-details information instead of transitions between relevant textual and pictorial information was one possible explanation for such a disruption of the learning process. Recent studies (Jian, Wu, & Su, 2014; Tsai, Hou, Lai, Liu, & Yang, 2012) already used an analysis of individual fixation sequences and comparisons of structural models of gaze direction to assess the eye movements during the construction of mechanical representations in more detail. Dewhurst, Nystöm, Jarodzka, Foulsham, Johansson, and Holmqvist (2012) also used different methods based on the “levenshtein distance” to compare series of gazes. Such an approach could be useful to assess the disruption hypothesis for seductive details within a multimedia learning instruction and to have a closer look on the structural or sequential component of information processing.

**Seductive-details effect due to distracting processes**

As the seductive-details group fixated the picture AOIs significantly later, this behavior can be assumed to be an indicator for a distraction effect. In combination with the moderating effects of prior knowledge these results give support for the distraction hypothesis (Sanchez & Wiley, 2006). Students with high prior knowledge were less affected by seductive details, which indicates with regard to the cognitive-load explanation that high prior knowledge learners experienced lower intrinsic cognitive load and more available capacity to compensate or even profit from seductive details (Park et al., 2011). Nevertheless, prior knowledge does not only reduce intrinsic cognitive load, it can also enhance the information selection. As an expertise related factor prior knowledge can enhance the ability to differentiate between relevant and irrelevant information to focus attention on relevant information processing and to decrease the seductive-details effect (Canham & Hegarty, 2010). Thus, the enhanced information selection could be another way how prior knowledge fosters to compensate for extraneous load effects like the seductive-details effect. In addition, Magner et al. (2014) found an interaction effect for seductive illustrations with prior knowledge and assumed that high prior knowledge learners are affected but only concerning high demanding cognitive activities that overburden their cognitive capacity. The results of the present study give support for this assumption. Within the used multimedia-learning instruction comprehension performance requires high demanding cognitive activity because complex mental models including moving parts have to be built out of pictorial and textual information. The results of the regression models indicate an increase in cognitive load also for high prior knowledge learners in the seductive-details condition, with a decrease in transitions and a decrease in learning success. However, the difference between high prior knowledge learners in the seductive-details group and high prior knowledge learners in the group without seductive details is not significant. Thus, seductive details also affected the high prior knowledge learners, but they were able to compensate in learning performance especially in contrast to the low prior knowledge learners in their own group. The rising question due to these results is now how learners with high prior knowledge compensate. Again further research and a more detailed analysis is needed to assess the question if high prior knowledge in contrast to low prior knowledge learners really tend to ignore seductive details and to investigate less cognitive activity in processing them or if they process the relevant information first and the seductive-details information afterwards.
Further support for the distraction hypothesis (Sanchez & Wiley, 2006) is given by the moderating influence of spatial ability. A study by Meneghetti, Gyselinck, Pazzaglia, and De Beni (2009), for example, shows the impact of spatial ability on mental model construction, processes of text-picture integration and its close link to working memory capacity. Thereafter, it is easier for learners with high spatial ability to construct mental models out of textual and corresponding pictorial information. In addition, high spatial ability learners are able to compensate under cognitively high demanding conditions and maintain their performance in contrast to low spatial ability learners. The present study gives support to these assumptions. The moderating effects for spatial ability show that high spatial ability learners were less affected by seductive details in comprehension even though they show comparable integrative transitions to the low spatial ability learners and less integrative transitions in contrast to the high spatial ability learners in the group without seductive details. Thus, transitions seem to be important especially for learners with low spatial ability and are indeed very important for information integration and mental model construction. As especially the information processing of the relevant pictorial information was affected by seductive details as well as the integrative transitions from textual to pictorial information, the assumption suggests itself that seductive details indeed affect cognitive processes of integration and model construction. In general it can be assumed that high spatial ability learners have more available capacities to process visuospatial information under cognitively high loading learning conditions so that they know to compensate seductive details and other extraneous cognitive-load effects (Seufert, Schütze & Brünken, 2009).

**Seductive-details effect due to perfunctory processes**

Concerning the measurement of cognitive load during learning activity the results in the subjective ratings (Paas, 1992) indicate lower cognitive load in the seductive-details group. This result contradicts to our hypothesis. A possible explanation for this result is that seductive details in the present study indeed decreased the germane cognitive load by introducing perfunctory information processing of the relevant information; therefore learners probably rated their cognitive activity based on that part of information processing. This assumption is supported by the decrease in total fixation duration and the total number of fixations on the relevant pictorial information, as well as by the decrease of integrative transitions between textual and pictorial information, that indicates less cognitive activity for relevant information processing (Just & Carpenter, 1976; Schmidt-Weigand et al., 2010). This perfunctory processing could also originate from diversion leading to the phenomenon of illusion of knowing (Glenberg, Wilkinson, & Epstein, 1982) in the way that learners seem to activate irrelevant schemata from the seductive-details material (examples of the use of ATP) instead of the relevant learning material (the structure and function of the ATP molecule) and thereafter have the illusion to have learned appropriately enough, however with low subjective-ratings in cognitive load due to their diverted focus of attention on the easy-to-learn material of seductive details. To answer this question further research is needed, especially concerning the interaction of the different cognitive-load factors. Further research should anyway involve objective cognitive-load measures like for example the rhythm method (Park & Brünken, 2015) and fine-grained instruments for subjective ratings to measure the different cognitive-load factors (Leppink, Paas, Van Gog, Van der Vleuten & Van Merrienboer, 2014).

**Final conclusion**

In sum, learners with low levels of spatial ability and prior knowledge were affected much stronger by seductive details than learners with high levels of both learner characteristics. The results show a decrease in cognitive load, indicated by perfunctory processing of the relevant information and a decrease in learning success. Information processing as well as learning success was moderated by spatial ability and especially by prior knowledge. Thus, the present study gives support for an individual approach concerning the research on learning and instruction with a focus on learner characteristics and shows that eye-movement analysis provides detailed information about the individual information processing that can be especially beneficial for the design of multimedia-learning instruction. The resulting practical implication for instructional designers, teachers and learners is only to use seductive details in case of already established high prior knowledge and/or in case of instructing high spatial ability learners. The most appropriate instructional design would be an adaptive learning-system where prior knowledge and spatial ability are tested before the learning session and learners profit from the individually adapted learning-environment due to the consideration of these different and most relevant learner characteristics in multimedia learning.
References


Gender Effects When Learning Manipulative Tasks From Instructional Animations and Static Presentations

Mona Wong1*, Juan C. Castro-Alonso1,2, Paul Ayres1 and Fred Paas1,3

1 School of Education, The University of New South Wales Sydney, Australia // 2Center for Advanced Research in Education (CIAE), Universidad de Chile, Chile // 3Institute of Psychology, Erasmus University, The Netherlands // p.wong@unsw.edu.au // jccastro@ciae.uchile.cl // p.ayres@unsw.edu.au // paas@fsw.eur.nl

*Corresponding author

ABSTRACT
Humans have an evolved embodied cognition that equips them to deal easily with the natural movements of object manipulations. Hence, learning a manipulative task is generally more effective when watching animations that show natural motions of the task, rather than equivalent static pictures. The present study was completed to explore this research domain further by investigating the impact of gender on static and animation presentations. In two experiments, university students were randomly assigned to either a static or animation condition and watched a computer-controlled presentation of a Lego shape being built. After each of two presentations, students were required to reconstruct the task followed by a transfer task. In Experiment 1 the tasks were performed using real Lego bricks (physical environment), and in Experiment 2 by computerized images of the bricks (virtual environment). Results indicated no differences between the two testing environments or an overall advantage for the animated format. However, a number of interactions between gender and presentation format were found. Follow-up analyses indicated that females benefited more than males from using animated presentations.

Keywords
Animation vs. static picture, Gender differences, Embodied cognition, Cognitive load theory, Technology-based learning

Introduction to the study

Are static or animated instructional presentations better for student learning? The answer to this question is rather complex. The research literature comparing instructional statics with animations provides different perspectives, challenging a single conclusion. There are studies supporting the use of animations over static images (e.g., Ardac & Akaygun, 2005; Lin & Atkinson, 2011; Ryoo & Linn, 2012), but there is also contrasting evidence suggesting that statics are superior to dynamic visualizations (e.g., Castro-Alonso, Ayres, & Paas, 2014b; Mayer, Hegarty, Mayer, & Campbell, 2005; Scheiter, Gerjets, & Catrambone, 2006). Moreover, there are comparisons showing no statistical differences between statics and animations (e.g., Mayer, DeLleuw, & Ayres, 2007; Narayan & Hegarty, 2002). There are even concerns about the validity of some of these comparisons (see Tversky, Morrison, & Bétrancourt, 2002). There are, as observed by Höffler and Leutner (2007), a number of important moderating variables affecting the instructional effectiveness of static and dynamic visualizations. One such moderator, revealed in the meta-analysis of 26 studies by Höffler and Leutner (2007), was the type of task to be learned: Animations and videos were most effective, as compared to statics, when procedural-motor tasks were depicted.

Evidence also shows the importance of spatial ability when learning from static or dynamic visualizations (see Höffler, 2010). However, there is a lack of consensus indicating whether high or low spatial ability correlates favorably when learning from dynamic images. Furthermore, highly related to spatial ability issues are gender effects. Research has generally found that spatial ability is higher in males (see Linn & Petersen, 1985; Uttal et al., 2013; Voyer, Voyer, & Bryden, 1995), but there is also evidence accumulating that animated presentations are particularly helpful for females rather than males (see Sánchez & Wiley, 2010; Yezierski & Birk, 2006), which supports the hypothesis that low spatial ability students (females) benefit most from animations. But again to muddy the waters in this field, there is also evidence suggesting that male students can outperform females in animated conditions (see Lin, Hung, Chang, & Hung, 2014).

In view of these inconsistent results more research is required in this domain to identify the conditions impacting on the effectiveness of animations. The main aim of the current study was to make such a contribution by conducting two experiments comparing males with females when learning about object manipulative tasks, using both animated
and static presentations. The following sections briefly outline the main theoretical aspects underpinning the investigation.

**Instructional animations of manipulative tasks**

As well as the meta-analysis of Höfﬂer and Leutner (2007), some more recent studies have shown that animations and videos that show human motor tasks are more effective than equivalent statics. This has been reported for motor tasks such as unscrambling puzzle rings (Ayres, Marcus, Chan, & Qian, 2009), copying origami paper designs (Wong et al., 2009), or constructing different knots (Garland & Sánchez, 2013; Marcus, Cleary, Wong, & Ayres, 2013). Thus, when human movement is involved, animations (and videos) seem to be more effective instructional tools than static pictures. This finding has been termed the **human movement effect** (see Paas & Sweller, 2012), and can be explained by the fact that we evolved cognitive mechanisms to imitate tasks involving human motion (see also Castro-Alonso, Ayres, & Paas, 2014a).

In particular, humans have evolved an **embodied cognition** that links cognition to the environment through the body (see Barsalou, 2010). Classical cognitive accounts (e.g., Atkinson & Shiffrin, 1968) tended to isolate the mental processes from the rest of the bodily activities. In contrast, the embodied account connects the processes of mind, body, and environment (see Wilson, 2002). For example, when manipulating an object with the hands, the mind’s mechanisms of observation and movement are linked with their corresponding body elements (eyes and hands) and with environmental cues (e.g., the object). The instructional implication of evolving an embodied cognition is that every visual learning task (perception) can be enhanced by bodily experiences (action). In consequence, the human movement effect is successful because it connects perception and action.

Restating, our cognitive system is wired towards linking perception to action. Arguably, the most important component to facilitate this connection is the **mirror neuron system** (also called the observation-execution matching system), composed of neurons that get activated both when observing and when imitating, for example, the manipulation of things (see Rizzolatti & Craighero, 2004). The first to describe this perception–action mechanism in humans were Fadiga, Fogassi, Pavesi, and Rizzolatti (1995), who found similar contractions in participants’ hand muscles when the subjects either observed other humans doing hand movements or performed directly these actions. Having this mirror system implies that, when we watch an object being moved by human hands, we are preparing to eventually manipulate it ourselves. In consequence, the mirror neuron system greatly facilitates the imitation and learning of object manipulations (see van Gog, Paas, Marcus, Ayres, & Sweller, 2009).

As a result of having a system that allocates resources to deal with manipulative tasks (besides many other actions), learning to manipulate objects is a relatively easy task for humans. Further, object manipulation can be classified under **biologically primary abilities** in the framework of **evolutionary educational psychology** (see Geary, 2007; Geary, 2008). According to Geary (1995) biologically primary abilities have evolved with our species, allowing humans to survive and develop in their natural environment. Understanding body language, and imitating object manipulations are primary abilities that have helped our species to survive by communicating and accessing essential natural resources (Geary, 2000). Because these abilities have evolved and thus improved over many millennia, the human cognitive system is prepared to use them very efficiently, with relatively low effort. In contrast, the majority of tasks included in formal instruction, **biologically secondary abilities**, have not evolved to the same degree, and thus they can be difficult to attain (Geary, 1995). In consequence, educational institutions could use the easier primary abilities as vehicles to teach the harder secondary abilities (see Geary, 2008; Sweller, 2008). The current study followed this approach by using object manipulation (primary ability) to teach a sequence that led to the construction of a final shape (secondary ability).

In summary, when learning object manipulations, animations may be more effective than static pictures, because we have evolved to learn these primary tasks following natural movement. The fact that animations may be effective does not imply that there are equally effective for everyone. For instance, the influence of spatial ability is very important for the effectiveness of manipulative animations and static pictures.
Instructional visualizations and spatial ability

When learning from an instructional animation, learners are processing visuospatial input. Processing visuospatial information involves a high cognitive load on working memory processors (cf. Ayres & Paas, 2007; Baddeley, 1996; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). This cognitive capacity to manage visuospatial information can be termed as spatial ability. Because it is employed with any visuospatial input, spatial ability is also used when learning from other types of instructional visualizations, such as static pictures.

A wide variety of evidence shows the importance of spatial ability when learning from static or dynamic visualizations (see Höffler, 2010). However, there is a lack of consensus on whether spatial ability is better to learn from either static or dynamic images. For example, the meta-analysis of 27 experiments by Höffler (2010) showed that high spatial ability students were more advantaged than low ability students when learning from statics rather than from animations. Similarly, Sánchez and Wiley (2014) showed that the spatial ability of psychology undergraduates was more significant in learning from text supplemented with static images rather than with animated presentations. These results correspond to the ability-as-compensator hypothesis described by Mayer and Sims (1994). This explanation states that spatial ability is required more when learning from static pictures, as spatial ability permits mental animation (Hegarty & Sims, 1994). The static depictions thus compensate for the lack of motion in the visualizations. In contrast, Yang, Andre, Greenbowe, and Tibell (2003) found that high spatial ability undergraduates scored higher on a transfer test when learning from animations as compared to statics. According to Mayer and Sims (1994) this result would be evidence for the ability-as-enhancer hypothesis, in which spatial ability enhances the benefits of animations over static pictures.

In conclusion, it is evident that spatial ability is a key factor in learning from visualizations. What is less clear is whether spatial ability is more helpful when watching either static or animated images, although there is more evidence supporting its effects on static images, as Höffler and Leutner (2011) commented. As a result of the central importance of spatial ability in processing instructional visualizations, students with low spatial ability will be disadvantaged. This unfavorable situation is generally observed in females when compared to males.

Gender, spatial ability and animations

Research tends to agree that spatial ability is higher in males than females, particularly for a spatial sub-dimension called mental rotation (see Linn & Petersen, 1985; Uttal et al., 2013; Voyer, Voyer, & Bryden, 1995). Greater outcomes for males in spatial tasks have been consistently reported in a variety of contexts (e.g., Campbell & Collaer, 2009; Collins & Kimura, 1997; Feng, Spence, & Pratt, 2007; Law, Pellegrino, & Hunt, 1993; Masters, 1998; Voyer & Hou, 2006). Regarding gender effects of spatial ability on instructional visualizations, the evidence is less male oriented.

For example, the inconclusive direction of the effects of spatial ability on static vs. dynamic images is similarly inconsistent when incorporating the gender variable. There are studies showing that instructional animations are an advantage for both females and males. For example, Yezierski and Birk (2006) assessed the effectiveness of an animated intervention on middle school, high school and university students. The authors observed that the pre-test scores favoring males disappeared after the animation treatment, showing that the dynamic images helped to close the initial gender gap. In other words, females learned more than males with the animations. Also, Sánchez and Wiley (2010) reported that psychology undergraduate males outperformed females in both a spatial ability test and in learning from text passages on a science topic. However, learning was similar between genders when the passages also contained animations, showing that these dynamic images were especially helpful for females and low spatial ability students.

This first group of studies support the ability-as-compensator hypothesis (Mayer & Sims, 1994), where low spatial ability participants (generally women), are benefited from the animated depictions, as they do not need to mentally animate these visualizations. Regarding a second group of studies, where animations favor high spatial ability students (generally men), Griffin, MacEachren, Hardisty, Steiner, and Li (2006) observed that males had better performance following animated presentations rather than statics, and that females performed equally under both visualization conditions. Also Lin, Hung, Chang, and Hung (2014) showed that male university students
outperformed females in the animated condition, but not in the condition presenting texts only. These two studies support the ability-as-enhancer hypothesis (Mayer & Sims, 1994).

As stated previously, it is not known which of the two hypotheses for spatial ability, either as compensator or as enhancer, is the most relevant when learning from instructional animations. Moreover, as far as we know, there are no studies addressing this issue for animations about object manipulative tasks. The present study aims to fill this gap in the literature.

**General description and hypotheses of the current study**

The current study compared the learning outcomes when imitating an object manipulative task after being modeled either in a static or animated presentation. In addition, gender effects were investigated. Consequently, a 2 (Presentation: static vs. animation) x 2 (Gender: male vs. female) between-subjects factorial design was used.

The manipulative task chosen was to complete a three-layered shape containing 15 Lego™ Duplo™ bricks of different lengths and colors based on the study by Castro-Alonso (2013). There was also a manipulative transfer task. In Experiment 1 all tasks were conducted physically, moving real Lego bricks with the hands. In contrast, in Experiment 2, the tasks were attempted in a virtual environment, moving Lego representations with the mouse. Previous results have shown that physical and virtual manipulations can yield equivalent outcomes (e.g., de Jong, Linn, & Zacharia, 2013; Zacharia & Olympiou, 2011). However, the identical elements theory (cf. Thorndike & Woodworth, 1901) and the congruence principle for effective graphics (Tversky et al., 2002) predict that doing the task in the physical environment may be more effective, as the presentations of this study involved real Lego bricks. Thus, by including two environments, we broadened the investigation, as well as providing opportunities for replication.

Two hypotheses were constructed. Because of the nature of the manipulative task a human movement effect was predicted in that animations would be superior to static presentations (Hypothesis 1). Castro-Alonso (2013) previously found support for this prediction with these same materials. The second hypothesis predicts a gender–presentation interaction. Even though research suggests that animation can favor both females and males, an interaction is likely considering the previous findings suggesting either males or females benefit most from animations. It was an open question as to which gender will benefit most in this study.

**Hypothesis 1:** Animations will be superior instructional materials to static pictures

**Hypothesis 2:** There will be a gender–presentation format interaction.

**Experiment 1**

The first experiment was conducted using a physical environment for testing.

**Method**

**Participants**

59 students (30 male, 29 female) aged between 17 and 40 ($M = 22.5, SD = 5.29$) were recruited from an urban Australian university. The sample consisted of 46 undergraduate students and 13 postgraduate students from various faculties including Arts and Social Sciences, Business, Engineering, Medicine and Science. They were randomly allocated into an animation group (16 male, 14 female) and a static pictures group (14 male, 15 female). Participants were given a $20 gift card for volunteering.
Materials

Background survey. This one-page questionnaire assessed gender, the university program, and the year that students were enrolled in, the level of study (i.e., postgraduate or undergraduate), and their handedness (right or left).

Assessment of spatial ability. To measure spatial ability, we employed the Card Rotations Test (Ekstrom, French, & Harman, 1976), which measures mental rotation of two-dimensional figures. Although the original test has two parts of 10 questions each, only the first part was used to maximize time for the experiment with a 3-minute completion allowance given.

Learning Materials. In the animation condition, a video depicting the 15 steps of a 3-D Lego construction (based on the materials used by Castro-Alonso, 2013) from an aerial view (see Figure 1) was filmed with a digital Sony Handycam in PAL standard (size 768 x 576 pixels; 25 frames per second) without audio. The video showed human hands (in green gloves) manipulating one Lego brick at a time, and placing them on a Lego platform one-by-one. The size of this animation was adjusted to a LEGO® platform’s size of approximately 200 x 200 pixels leaving the Lego platform placed at the center of the screen filled with white background. The video was then edited and exported with Adobe Premiere Pro CS3 (Adobe, 2007) to Adobe Flash Video format (.flv). In total, the animation lasted 92 seconds.

The video was resized to 200 x 200 pixel so that it was the same size as a single picture form the static picture condition (see Figure 1 left). Furthermore, a numbering system indicating the brick sequence was added onto the video to match with that in the static picture condition.

![Figure 1. Experiment 1 animation condition (left) and static-picture condition (right)](image)

In the static picture condition, 15 key frames showing each brick being placed was extracted from the animation. Each key frame had identical size and information as in the corresponding video (see Figure 1 right). All 15 frames, sequentially numbered on the top left, were presented simultaneously on the screen and could be viewed for the same time as in the animation presentation (92 seconds). It was important that every significant change (each brick placement) in the animation was also shown in the static condition, and that both presentation formats were sized identically. Extensive pilot testing of these materials suggested that no bias existed in favor of animation. In this fashion the concern expressed by Tversky et al. (2002) that animations often contain more information than statics was eliminated as much as possible.

For both animation and static conditions, participants had no control over the pacing of the learning materials.

Testing Environment

Completion Task. The same set of Lego Duplo bricks used in filming the learning material was used for testing. In this physical environment the actual bricks were given to the participants, arranged in a vertical position, according to the order that they appeared in the learning materials starting from top left to bottom right (see Figure 2). A brown square building platform was provided on their work desk for participants to build the required shape on. This platform had a fixed orientation identical to the learning presentation. Participants were required to build the shape they viewed in the presentation. They were told that the same set of bricks, in the same order, as shown in the
presentation were given, and only one brick could be placed at a time starting from the top left to right bottom (See Figure 1). No changes were allowed once the previous brick position was confirmed and the next brick was picked.

![Figure 2. Performing environment used in completion task for Experiment 1](image)

Transfer Task. The first six red Lego bricks, which were used in the completion task were reused in the transfer task. All 6 bricks were placed vertically and were arranged according to the order in the completion task starting from left to right (the six red bricks in the top line of Figure 2). Similar to the completion task, the square platform was provided as a basis for the building. For the transfer task participants were required to recall only the bottom layer (i.e., the first 6 bricks) from the learning material, but rebuild them in a 90-degree clockwise rotation. Participants were explicitly told not to rotate their heads nor the platform (which was fixed), instead, they were required to complete the rotation mentally.

Grading rubric. To ensure precise scoring of the tasks, a detailed grading rubric was developed. In completion tasks 1 and 2, one mark was given to each brick if the brick (same shape and color) was placed in its correct position regardless of the order in which it was placed. However, 1 mark was deducted if a) two or more adjacent bricks were placed in the correct position but were rotated; b) the bricks were place in the correct position and orientation, but the brick positions were switched; c) both orientation and configuration were correctly recalled, but the whole structure was shifted away from its correct position. Each level was scored separately – if a brick was placed in its correct position but on a different level, then no marks would be given for that brick. For the two completion tasks, the maximum score was 15 and the minimum was 0. In the transfer task, all the above rules applied. Additionally, 1 extra mark was given to those building with the correct (90 degree clockwise) correct configuration, sequence and orientation. The maximum score for the transfer task was 7 and the minimum was 0.

Self-report of cognitive load. To get a measure of cognitive load, a self-rating measure was used based on the subjective scale of Paas (see Paas, 1992; Paas & van Merriënboer, 1994). Participants were asked how much mental effort they spent in completing the task right after they completed each task. The responses were given on a 9-point Likert scale ranging from little (1) to fair (5) to heavy (9).

Procedure

The experiment sessions were conducted in a quiet room. Each session lasted about 40 minutes, and only one student was tested in each session. After completing the survey, the participants attempted the Card Rotations Test for 3 minutes. After a practice task, the participants watched the assigned learning task materials for the first time. Immediately afterwards they were required to build the shape (the 1st attempt). During this attempt participants had no access to the learning materials. Immediately after completion of their construction, they were required to rate their mental effort. This was then followed by a repetition of the procedure by watching the learning task for a second time and then attempting the construction (2nd attempt), and completing the mental effort scale. Immediately
after the 2nd attempt of the main task, participants were given the transfer task followed by the mental effort rating of the transfer task.

Results

Tests of ANOVA assumptions

To test if the assumptions for using ANOVA were met, individual group normality tests (Kolmogorov-Smirnov) and Levene’s test for Homogeneity of variance were completed for all measures. The results indicated that all assumptions were met, and therefore ANOVA was used throughout this experiment.

Test of spatial differences

To test for initial spatial differences amongst the groups, the spatial measures (CRT test) collected before the acquisition phase were analyzed using a 2 x 2 ANOVA. The spatial measure (CRT scores) was not significant for gender ($F < 1$, $ns$), presentation type ($F < 1$, $ns$), or interaction ($F < 1$, $ns$). Hence, the CRT spatial measure was not used as a covariate and 2 (animation vs. static picture) × 2 (male vs. female) ANOVAs were used to investigate the hypotheses in this study.

Test scores

Group mean scores for each test (reconstruction attempts and the transfer test) score are reported in Table 1. Maximum scores for the 1st and 2nd attempts were 15, and the transfer test was 7.

<table>
<thead>
<tr>
<th></th>
<th>Animation</th>
<th>Static picture</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>5.43 (3.52)</td>
<td>6.14 (3.44)</td>
<td>5.77 (3.44)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>7.22 (4.55)</td>
<td>9.82 (4.51)</td>
<td>8.43 (4.64)</td>
</tr>
<tr>
<td>Transfer</td>
<td>3.88 (2.60)</td>
<td>4.07 (2.75)</td>
<td>3.97 (2.63)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>7.00 (2.50)</td>
<td>5.43 (3.08)</td>
<td>6.19 (2.88)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>10.75 (2.83)</td>
<td>8.80 (4.10)</td>
<td>9.74 (3.62)</td>
</tr>
<tr>
<td>Transfer</td>
<td>5.18 (1.58)</td>
<td>3.63 (2.64)</td>
<td>4.38 (2.29)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>6.17 (3.13)</td>
<td>5.78 (3.22)</td>
<td>5.98 (3.16)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>8.87 (4.19)</td>
<td>9.29 (4.25)</td>
<td>9.08 (4.19)</td>
</tr>
<tr>
<td>Transfer</td>
<td>4.48 (2.25)</td>
<td>3.85 (2.65)</td>
<td>4.17 (2.46)</td>
</tr>
</tbody>
</table>

Figure 3. Completion interaction for the second attempt in Experiment 1
Two-way ANOVAs showed that there was no significant main animation effect for the 1st attempt \( (F < 1, \, ns) \); 2nd attempt \( (F < 1, \, ns) \); or the transfer task, \( F(1,55) = 1.12, \, p = .30, \, \eta^2 = .02, \, MSE = 5.20 \). In addition, there was no significant main gender effect for the 1st attempt, \( (F < 1, \, ns) \); the 2nd attempt, \( F(1,55) = 1.39, \, p = .24, \, \eta^2 = .03 \); or for the transfer task, \( F < 1, ns \). There was no significant interaction for the 1st attempt, \( F(1,55) = 1.12, \, \eta^2 = .02 \), \( p = .30 \), or the transfer task \( F(1,55) = 1.86, \, \eta^2 = .03 \). However, there was a significant interaction for the 2nd attempt, \( F(1,55) = 4.59, \, \eta^2 = .08 \) (see Figure 3). Follow-up simple effects tests indicated that for the animation format, females \( (M = 10.75) \) scored significantly higher than males \( (M = 7.22) \); \( F(1,28) = 6.28, \, \eta^2 = .18 \). However for the static format, there was no significant gender difference \( (F < 1, ns) \).

Cognitive load scores

Group means scores for each cognitive load measure are reported in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Animation Mean (SD)</th>
<th>Static picture Mean (SD)</th>
<th>Total Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>7.88 (1.5)</td>
<td>6.79 (1.93)</td>
<td>7.37 (1.77)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>7.25 (1.65)</td>
<td>6.57 (1.83)</td>
<td>6.93 (1.74)</td>
</tr>
<tr>
<td>Transfer</td>
<td>5.81 (2.23)</td>
<td>5.14 (2.48)</td>
<td>5.5 (2.33)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>8.00 (0.88)</td>
<td>7.13 (1.25)</td>
<td>7.55 (1.15)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>7.50 (1.16)</td>
<td>6.73 (1.44)</td>
<td>7.10 (1.35)</td>
</tr>
<tr>
<td>Transfer</td>
<td>6.93 (1.82)</td>
<td>6.47 (2.42)</td>
<td>6.69 (2.12)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>7.93 (1.23)</td>
<td>6.97 (1.59)</td>
<td>7.46 (1.49)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>7.37 (1.43)</td>
<td>6.66 (1.44)</td>
<td>7.02 (1.55)</td>
</tr>
<tr>
<td>Transfer</td>
<td>6.33 (2.09)</td>
<td>5.83 (2.49)</td>
<td>6.08 (2.29)</td>
</tr>
</tbody>
</table>

Results indicated a significant main effect for presentation format for the 1st attempt, \( F(1,55) = 6.80, \, \eta^2 = .11, \, MSE = 2.07 \); where the participants in animation group \( (M = 7.94) \) rated cognitive load higher than in the static picture group \( (M = 6.96) \); a close to significance \( (p < .10) \) for the 2nd attempt, \( F(1,55) = 3.23, \, \eta^2 = .06, \, MSE = 2.38 \); where again the animation group \( (M = 7.38) \) reported higher cognitive load than static picture group \( (M = 6.65) \), but no significance for the transfer task \( (F < 1, ns) \). There were no significant gender effects at the first attempt \( (F < 1, ns) \), or the second attempt \( (F < 1, ns) \). For the transfer task, there was a significant gender effect, \( F(1,55) = 4.32, \, \eta^2 = .07 \), where females reported spending higher cognitive load \( (M = 6.69) \) than males \( (M = 5.47) \). There were no significant interactions (all \( F < 1, ns \)).

Discussion

There was no support for Hypothesis 1 that predicted an animation effect, as no evidence was found that animations were superior to statics on this task. The only significant effect was found for the cognitive load measure on the first attempt, where more cognitive load was recorded in the animation condition. A similar non-significant \( (p < .10) \) result was found at the second attempt. These results suggest that learning through animations required greater cognitive load, but this did not directly impact on performance. On the transfer task females reported higher mental effort than males, which again did not impact on learning.

Hypothesis 2 predicted a gender–presentation format interaction. On test scores there was a significant interaction for the second attempt at the task. Simple effects tests indicated that for the animation format females outperformed males, but for the static format no significant gender effects were found. Although, no other significant interactions were found for test scores, examination of Table 2 indicates that for both the first attempt and the transfer task, females had higher scores than males for the animated condition, but this pattern was not repeated for static presentations.
Experiment 2

As outlined in the introduction, Experiment 2 replicated the experimental design of Experiment 1 using a virtual testing platform instead of a physical testing platform. The same two hypotheses were again tested.

Method

Participants

86 students were recruited from a large university in Sydney (42 male, 44 female) participated in the experiment. They were enrolled in courses from various faculties including Arts and Social Sciences, Business, Engineering, Medicine and Science. The sample consisted of 72 undergraduate students and 14 postgraduate students, aged between 17 and 46 (M = 21.9, SD = 5.64). They were randomly allocated into four groups: animation (22 male, 22 female) and static pictures (20 male, 22 female).

Materials

The same learning materials and tasks from Experiment 1 were re-employed in Experiment 2. The only, and major, difference was the testing environment that participants built the Lego pattern onto. In contrast to Experiment 1, participants were required to rebuild the assigned pattern on a computer (virtual platform). The virtual platform (see Figure 4) was developed on Actionscript 3 with Adobe Flash CS4 Professional (Adobe, 2008). The same number and color of the bricks used in the learning and testing materials in Experiment 1 were created virtually. The participants were required to move the bricks and rebuild the Lego shape onto the square building-stage (shown in the top left of Figure 4). To reposition the bricks participants had to drag the bricks with the computer mouse, and double-click to rotate the bricks. Only one brick could be moved at a time. After participants had placed a brick, the next button needed to be clicked (shown in the top right of Figure 4) in order to move forward to the next brick. Moreover, once this button was clicked, no further modifications were possible. For the transfer task, the first six bricks used in the completion task were required. All six bricks were presented virtually and in identical order to the completion task starting from left to right.

Procedure

The same procedures and times (learning and testing) used in Experiment 1 were again followed. The only difference was that instead of the experimenter bringing the real physical bricks to the participants after they finished watching the learning materials, the building platform appeared automatically on the computer screen. There were three testing tasks: completion task 1 (1st attempt), completion task 2 (2nd attempt) and a transfer task. Cognitive load (mental effort) measures were collected after each task completion.
Results

Tests of ANOVA assumptions

Similar to Experiment 1 the Kolmogorov-Smirnov and Levene’s tests were used to check that the assumptions for using ANOVA were met. All assumptions for the first and second completions were met; however, for the transfer task, the distribution of male scores in the animated format failed the normality test, $D(22) = .24, p = .002$, as did females in the animated format, $D(22) = .21, p = .011$. Furthermore, male transfer scores in the static format also failed the normality test, $D(20) = .39, p < .001$. There was a high proportion of students who achieved maximum scores, thus skewing the data. Hence, ANOVAs were used in the 1st and 2nd attempt of the completion tasks, and non-parametric methods were used in the transfer task analysis.

Test of spatial differences

To test for initial spatial differences amongst the groups, the spatial measures (CRT test) collected before the acquisition phase were analyzed using a 2 x 2 ANOVA. There was no main effect for gender, $F(1, 82) = 2.11, p = .15, \eta^2_p = .03$; presentation type, $F(1,82) = 2.90, p = .09, \eta^2_p = .03$; or a significant interaction, $F(1,82) = 1.11, p = .29, \eta^2_p = .01$. Hence, the CRT spatial measure was not used as a covariate, and 2 (animation vs. static picture) × 2 (male vs. female) ANOVAs were again used to investigate the hypotheses in this study.

Test scores

Group mean scores for each test (reconstruction attempts and the transfer test) score are reported in Table 3. Maximum scores for the 1st and 2nd attempts were 15, and the transfer test was 7.

<table>
<thead>
<tr>
<th></th>
<th>Animation</th>
<th>Static picture</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>6.23 (3.25)</td>
<td>6.63 (2.95)</td>
<td>6.42 (3.08)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>9.84 (4.07)</td>
<td>11.33 (3.21)</td>
<td>10.55 (3.72)</td>
</tr>
<tr>
<td>Transfer</td>
<td>5.00 (2.12)</td>
<td>5.65 (2.12)</td>
<td>5.31 (2.12)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>7.68 (1.82)</td>
<td>6.32 (2.55)</td>
<td>7.00 (2.30)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>11.36 (2.30)</td>
<td>9.91 (3.48)</td>
<td>10.64 (3.00)</td>
</tr>
<tr>
<td>Transfer</td>
<td>5.80 (1.41)</td>
<td>4.64 (2.13)</td>
<td>5.22 (1.88)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>6.96 (2.70)</td>
<td>6.46 (2.72)</td>
<td>6.72 (2.71)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>10.60 (3.35)</td>
<td>10.58 (3.39)</td>
<td>10.59 (3.35)</td>
</tr>
<tr>
<td>Transfer</td>
<td>5.40 (1.83)</td>
<td>5.12 (2.16)</td>
<td>5.26 (1.99)</td>
</tr>
</tbody>
</table>

The 2 (animation vs. static picture) × 2 (male vs. female) ANOVAs showed that there was no significant main animation effect for the 1st attempt ($F < 1, ns$) and the 2nd attempt ($F < 1, ns$). Likewise, there was also no significant main gender effect for the 1st attempt ($F < 1, ns$), and the 2nd attempt, ($F < 1, ns$). Also, there was no significant interaction for the 1st attempt $F(1, 82) = 2.30, p = .13, \eta^2_p = .03$.

However, there were significant interactions for the 2nd attempt $F(1, 82) = 4.18, p < .05, \eta^2_p = .05$ (see Figure 5). Follow-up simple effect tests showed no significant gender differences for the 2nd attempt in the animation format, $F(1, 42) = 2.34, p = .13, \eta^2_p = .05, MSe = 10.91$, or the static format, $F(1, 42) = 1.87, p = .18, \eta^2_p = .05, MSe = 11.25$.

Based on the previous interactions found, Mann-Whitney tests were conducted to examine potential gender simple effects. For animation, no significant differences was found between genders, $U(44) = 201.5, Z = -.99, p = .32, r = -.15$. However, for the static format, males (mean rank = 24.9) scored higher (significant at $p < .10$) than females (mean rank =18.41), $U(42) = 152.0, Z = -1.80, p = .07, r = -.28$. 

46
Cognitive load measure

Group mean scores for each cognitive load measure are reported in Table 4 respectively below.

Table 4. Mean (SD) for cognitive load measure in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Animation</th>
<th>Static picture</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>7.77 (1.51)</td>
<td>7.10 (1.29)</td>
<td>7.45 (1.44)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>6.91 (1.57)</td>
<td>6.25 (1.74)</td>
<td>6.60 (1.67)</td>
</tr>
<tr>
<td>Transfer</td>
<td>6.09 (1.85)</td>
<td>5.15 (2.37)</td>
<td>5.64 (2.14)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>7.68 (1.00)</td>
<td>7.68 (1.09)</td>
<td>7.68 (1.03)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>6.77 (1.11)</td>
<td>6.73 (1.52)</td>
<td>6.75 (1.31)</td>
</tr>
<tr>
<td>Transfer</td>
<td>6.18 (2.26)</td>
<td>6.00 (2.05)</td>
<td>6.09 (2.13)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st attempt</td>
<td>7.73 (1.26)</td>
<td>7.40 (1.21)</td>
<td>7.57 (1.24)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>6.84 (1.35)</td>
<td>6.50 (1.63)</td>
<td>6.67 (1.49)</td>
</tr>
<tr>
<td>Transfer</td>
<td>6.14 (2.04)</td>
<td>5.60 (2.22)</td>
<td>5.87 (2.14)</td>
</tr>
</tbody>
</table>

Results indicated there was no significant main animation effect for the 1st attempt $F(1, 82) = 1.59, p = .21, \eta_p^2 = .02, MSe = 1.53$; for the 2nd attempt, $F(1, 82) = 1.19, p = .28, \eta_p^2 = .01, MSe = 2.24$; and for the transfer task, $F(1, 82) = 1.45, p = .23, \eta_p^2 = .02, MSe = 4.56$. Also, there was no significant main effect for gender for the 1st attempt ($F < 1, ns$), for the 2nd attempt ($F < 1, ns$), and for the transfer task, $F(1, 82) = 1.04, p = .31, \eta_p^2 = .01$. There was also no interaction for the 1st attempt, $F(1, 82) = 1.59, p = .21, \eta_p^2 = .02$; for the 2nd attempt ($F < 1, ns$), and for the transfer task ($F < 1, ns$).

Discussion

Hypothesis 1 predicted an animation effect and consistent with Experiment 1 no overall advantage was found for the animated format. However, support was found for Hypothesis 2 in that an interaction was found on the second attempt indicating that comparative performance between males and females was moderated by the original presentational format. Although the simple effect tests were not significant, examination of Table 3 indicates that again, females achieved higher scores than males for the animation condition, and males scored higher than females for the static condition, results consistent with the pattern identified in Experiment 1.

General discussion

A significant presentation–gender interaction was found in both experiments for this object manipulative task (Lego construction). Follow up simple effects tests indicated that for the second attempt at the completion test in Experiment 1 females scored significantly higher than males in the animated format. In contrast, this female advantage was not found for static presentations. Overall, the pattern of results was very consistent. As can be seen
in Tables 1 and 3, for every test females had higher scores than males in the animated condition, but for the static condition males had higher scores than females, suggesting a clear pattern. The evidence suggests that in this learning domain with these materials, gender is an important moderator. Some caution needs to be shown in generalizing results from a single study; however, the research may be significant for a number of reasons.

Firstly, the results are consistent with previous research on gender and animation that found that animation helps female learning. However, unlike previous research that has focused mainly on learning topics related to science (Falvo & Suits, 2009; Jacek, 1997; Sánchez & Willey, 2010; Žeziški & Birk, 2006), we conducted our study with an objective manipulative task. Investigating such a task is important, because as previously argued in our theoretical introduction, animations may be more suited to tasks involving human motion as we have evolved cognitive mechanisms to imitate such tasks (Castro-Alonso et al., 2014a). In this study, it is potentially an important finding that females responded more favorably than males, on a task that was predicted to produce an overall animation effect (Hypothesis 1) similar to that found by Castro-Alonso (2013). It is notable that in the Castro-Alonso study, consistent with much research using Education and Psychology students, there were significantly more female participants than males. If the gender effect found in this study can be generalized then it is possible that previous studies that included high percentages of females may have been biased. Such a conclusion needs further investigation.

Secondly, the testing in Experiment 1 was conducted in a physical environment, whereas Experiment 2 was conducted on a virtual platform. The results from both experiments were fairly consistent suggesting that the type of environment did not impact on the presentation format or gender. These results are in line with other studies that have found no difference between virtual and physical environments (e.g., de Jong, Linn, & Zacharia, 2013; Klahr, Triona, & Williams, 2007; Zacharia & Olympiou, 2011), and no advantage was found for the physical environment due to the congruence principle for effective graphics (Tversky et al., 2002) or the identical elements theory (cf. Thorndike & Woodworth, 1901). Crucially gender effects did not seem to be moderated by testing format either, although it should be noted that environments were not compared directly in the same experiment.

Thirdly, together with the research indicating that males have higher spatial ability than females (e.g., Campbell & Collaer, 2009; Collins & Kimura, 1997; Feng, Spence, & Pratt, 2007; Law, Pellegrino, & Hunt, 1993; Masters, 1998; Voyer & Hou, 2006), researchers have argued that animation helps low-spatial ability learners, in particular females. However, our study measured spatial ability using the Card Rotational Test (CRT) and no gender difference was found on this test, suggesting that spatial ability could not explain the interactions found. This conclusion raises two issues. Is the Card Rotational Test a good test in relation to an animation-presentation format or motor task learning? Although used extensively as a measure of spatial ability it was designed specifically to measure mental rotation, which may not be the best index for mental animation, although the transfer test in our study required mental rotation. Further research using more extensive spatial ability tests for animation studies may be needed in future to fully explore this domain. On the other hand, if the CRT was a reliable test, then it is an open question as to what other factors may have caused the gender effects found in this study. Some researchers have argued that females might have better visual recognition memory than males (e.g., McGivern et al., 1998), which may give them an advantage in animated designs. However, it might be expected that static pictures would also benefit from enhanced visual recognition memory. Clearly more research is needed to understand why females might be advantaged by animations.

From the perspective of managing cognitive load (the theme for this special issue), the use of instructional animations is not necessarily a panacea for more effective learning. As described in the literature review, animations can create transitory information that makes learning harder than from static presentations, unless human movement skills are involved. The effectiveness of animations is dependent upon many factors such as learning content and gender. Under many conditions cognitive load will be raised using animations; hence use of animations needs to be carefully matched with learners and content to manage cognitive load. The evidence collected in this study suggests that for females, on such tasks as investigated here, using animations may have clearer advantages in managing their cognitive load rather than statics. For males, the reverse strategy may be more effective.

In summary, the main aim of the current study was to make a theoretical and empirical contribution to this general domain by comparing males with females when learning about object manipulative tasks, using both animated and
static presentations. We found that for an object manipulative task (Lego construction), instructional format interacted with gender. For instructional animation, some evidence was found that university females outperformed males. However, for a static presentation, no gender differences were found. It can be cautiously concluded, that animations enable females to learn such tasks better, but for males an animation presentation may be redundant (see Chandler & Sweller, 1991) as they can at least learn equally well from static presentations. The next step in the process is to replicate these results through further research using different spatial ability tests and different learning domains.

Acknowledgments

This research was partially supported by funds from CONICYT (PAI, national funding research program for returning researchers from abroad, 2014, No 82140021; and Basal Funds for Centers of Excellence, Project FB0003) to the second author.

References


The Impact of Supported and Annotated Mobile Learning on Achievement and Cognitive Load

Rustam Shadiev¹, Wu-Yuin Hwang², Yueh-Min Huang* and Tzu-Yu Liu³
¹Department of Engineering Science, National Cheng Kung University, Taiwan // ²Graduate Institute of Network Learning Technology, National Central University, Taiwan // ³Bei-Zheng Junior High School, Taiwan // rustamsh@gmail.com // wyhwang@cc.ncu.edu.tw // huang@mail.ncku.edu.tw // dtyliu@gmail.com
*Corresponding author

ABSTRACT
We designed activities for learning English as a foreign language in a mobile learning environment with familiar authentic support for this study. Students learned at school and then applied their newly gained knowledge to solve daily life problems by first using a tablet to take pictures of objects they wished to learn about, then describing them and sharing their homework with peers. For this study two experiments were carried out in which 59 junior high school students participated. A class of 28 students served as the control group in Experiment 1, and as the experimental group in Experiment 2; a second class of 31 students served as the experimental group in Experiment 1 and as the control group in Experiment 2. In the class serving as the control group, students studied and completed each learning activity using traditional textbooks while the experimental group studied using an electronic textbook and used a learning system installed on tablet PCs. This study investigates the effects of the mobile system on learning achievement and cognitive load. The research resulted in three main findings. First, the experimental students outperformed the control students on post-test items in both experiments. Second, learning activities using the tablet learning system caused less cognitive load for the students than when learning without technological support. Finally, this study found that creating text annotations is very important learning behavior and it predicts learning achievement. Based on these results, several implications along with conclusions and suggestions for future research are suggested at the end of this study.

Keywords
Mobile learning, Familiar authentic environment, Annotation, Learning achievement, Cognitive load

Introduction
The rapid advancement of information and communication technology has created new opportunities for curriculum design (Liang & Huang, 2014; Shadiev, Hwang, & Huang, 2015). For example, according to a growing number of studies, mobile technology provides advantages such as giving the ability to learn anywhere and at any time (Hwang, Chen, Shadiev, Huang, & Chen, 2014). Chang, Tseng, and Tseng (2011), Chu (2014), and Hwang, Shih, Ma, Shadiev, and Chen (2015) suggest that mobile technology creates an authentic learning environment in a real-world context in which learners adapt learning content to the context they find themselves in. Furthermore, mobile technology can be used as a cognitive tool to aid learning by decreasing learners’ cognitive load (Hwang, Wu, Zhuang, & Huang, 2013). Following these discoveries, mobile-assisted learning has been successfully implemented in a number of studies in various disciplines to manage cognitive load and facilitate learning; for example, in a social science course (Hwang et al., 2013; Chu, 2014) and in foreign language learning (Chen, Hsieh, & Kinshuk, 2008; Chang et al., 2011).

However, in spite of the potential advantages of this approach, the present study has also uncovered some potential issues that have not yet been addressed appropriately in related research. For example, previous studies have not dealt with developing speaking and writing skills or the ways in which knowledge is applied. In addition, most related studies have only focused on the effects of mobile learning on cognitive load in general; being in a familiar context in an authentic environment was not yet considered. Therefore, this study seeks to remedy these problems through the design of various speaking and writing learning activities to be carried out in a familiar authentic environment supported by mobile technology.
Literature review

Cognitive load

Mayer and Moreno (2003) and Plass, Chun, Mayer, and Leutner (2003) argue that taking cognitive load into consideration should be a central principle in the design of effective instruction because working memory limits learners’ cognitive capacity to accommodate demands imposed by learning tasks. According to Mayer and Moreno (2003) and Plaas et al. (2003), learning performance can be negatively affected when cognitive load exceeds the limit of cognitive capacity. Thus the issue of how to reduce cognitive capacity overload while encouraging knowledge construction should be considered when attempting to improve instruction design.

Three types of cognitive load are distinguished in the literature: intrinsic, extraneous, and germane (Brunken et al., 2003; Sweller et al., 1998). Intrinsic load is determined by the inherent nature of the learning material, learners’ expertise and an interaction between them; that is, the amount of information units that a learner needs to hold in working memory to comprehend the information. It is argued that intrinsic load is not affected by the instructional design but only by the learning material. Therefore, to avoid exceeding the appropriate intrinsic load, instructors need to adjust learning materials to meet the learners’ expertise. The term “extraneous load” refers to the cognitive load caused by the format and the manner in which information is presented as well as by the working memory requirements required to perform the actual instructional activities. An extraneous load can be imposed by improper instructional design. Thus, to keep the extraneous load from becoming excessive, instructors need to organize, present and carry out learning information and activities appropriately. A germane load is determined by learners’ efforts to process and comprehend the learning material. This load is also associated with motivation and interest. A germane load is induced by appropriate instructional design and can enhance learning.

Cognitive load in foreign language learning

Numerous studies on foreign language learning have investigated how to keep learners from being cognitively overloaded. Some information processes in foreign language learning are complex and impose a heavy cognitive load on working memory (Keysar, Hayakawa, & An, 2012). For example, applying learned vocabulary to new contexts involves high interactivity of elements, which causes a high intrinsic load (Plass et al., 2003). To reach the intermediate level of a foreign language, thousands of words need to be learned (Hulstijn & Laufer, 2001). When learners apply learned vocabulary to a new context, they have to focus on output (i.e., what to say/write and how to say/write it) which requires that several language forms (i.e., vocabulary and sentence patterns) be retrieved from long-term memory (Nation, 2003). Another example is when learners listen to a lecturer to comprehend instructional content; learners need to retain information in working memory integrate it with what follows, and all the while continually adjust understanding of incoming information with prior knowledge (Chen & Chang, 2009).

Several approaches have been proposed to develop instructional methods to teach learners how to efficiently apply what they have learned to new contexts using their limited working memory capacity. Of particular interest to this study are those approaches based on multimedia support. Cognitive theory of multimedia learning postulates that people learn more when information is presented through more than one media, e.g., words and pictures (Mayer, 2009). This is because learners process the information through multiple channels, e.g., auditory and visual; consequently, learning is deeper and information is retained in the memory longer. Acha (2009) and Plass et al. (2003) suggest that different presentation modes (i.e., visual and verbal) can ease the information processing load. Many tasks have visual and verbal representations; therefore, learners can easily access them, keep less learning information in working memory, and use more attentional resources elsewhere (Skehan, Willis, & Willis, 1996).

Children in a study performed by Acha (2009) read a short story in English on a computer screen and received verbal annotations (written translation). Acha (2009) found that verbal annotation affects vocabulary learning as children establish a direct connection between the verbal representation of the word in the native language and its foreign equivalent. English-speaking students of Plass et al. (2003) read a story in German on a computer and received verbal annotation, visual annotation, or neither. It was found that reading comprehension was worse with only the visual-annotations than with either no annotations or both verbal and visual annotations. Apparently the visual image annotation, when presented alone, introduce confusion, especially for words that are difficult to depict visually.
Learning in authentic environment with multimedia support

Context plays an important role in learning (Hwang et al., 2014a; Shadiev et al., 2015). Therefore, when designing learning activities, educators have to ensure that the environment where the learning activity takes place is authentic, i.e., relevant to real-life situations. According to Hwang et al. (2014a), an authentic environment (1) provides contexts that reflect the way the knowledge will be used in real life, (2) provides authentic activities that have real-world relevance, and (3) creates an opportunity for learners to share their learning and accessing experiences with various levels of expertise.

According to Chang et al. (2011), Chu (2014), and Huang and Chiu (2014), mobile technology can be used in authentic contexts with rich resources for students to learn. Furthermore, mobile-assisted learning offers a seamless learning experience, i.e., it can be accessed at any time or place (Hwang et al., 2014a). The effectiveness of multimedia-based instruction has been emphasized in many previous studies. With mobile multimedia tools, students can create learning materials in an authentic environment (Huang, Huang, & Wu, 2014). Multimedia aids, such as pictures and audio, make learning more interactive and information rich (Huang et al., 2011). Moreover, multimedia objects in learning stimulate students’ imagination and help bring out meaningful output (Huang, 2014). Multimedia aids, such as pictures and audio, make learning more interactive and information rich (Huang et al., 2011). Moreover, multimedia objects in learning stimulate students’ imagination and help bring out meaningful output (Caldwell, 1998). Students in the study of Hwang et al. (2014c) took pictures of learning objects in authentic contexts and then described them by using vocabulary and grammar learned in class. In the study of Hwang, Shadiev, Wu, and Chen (2014b), students practiced the target language by speaking out learning materials from the textbook, taking pictures of learning objects from daily life, and orally introducing them. Hwang, Shadiev, and Huang (2011) argue that, with multimedia aids, students can practice the target language repeatedly and regularly and expose themselves to diverse learning objects which increase the richness of their language experience. Harmer (2007) suggests that if students record their speech they (and the instructor) can listen to their own recordings, evaluate language performance, and see how much progress they have been making. Without multimedia support students need to hold a mental representation of the context in working memory over a period of time which is called “representational holding” (Mayer & Moreno, 2003). According to Mayer and Moreno (2003), when students attempt to engage in both information processing (i.e., selecting, organizing, and integrating learning material) and representational holding, cognitive overload occurs. Distributed cognition theory suggests that students’ representational holdings can be transformed into artefacts (i.e., pictures and their descriptions in written or oral form) which may prevent students being cognitively overloaded (Hollan, Hutchins, & Kirsh, 2000). Hollan and his colleagues argue that knowledge and cognition are not confined to an individual but distributed by placing memories, facts, or knowledge of the objects, individuals, and tools in the learning environment as a set of representations.

Hwang et al. (2011) suggest that the learning material and relevant multimedia aids should be placed on the same screen. In this way, a connection can be built between learning materials and supporting multimedia that gives students a clear picture of the whole learning scenario. Mayer and Moreno (2003) call this form of presentation “integrated presentation.” Using this approach, learners are more focused on essential information processing; that is, more cognitive capacity can be activated.

Huang et al. (2011) claim that multimedia supports students in their efforts to communicate in the target language with less anxiety about making mistakes. Chen et al. (2009) asserts that anxiety is a subjective feeling of worry, nervousness, or unease associated with arousal of the autonomic nervous system. Anxiety interferes with the cognitive ability to absorb, process, and produce a foreign language. Furthermore, it negatively affects cognitive load.

Hwang et al. (2011) claim that sharing homework with peers allows further reflection, discussion, and collaboration. In addition, sharing homework increases practice opportunities and helps students engage in English as Foreign Language (EFL) contexts. For example, students can listen to others’ audio recordings in which they can hear a diversity of speeches (i.e., variations in accent, fluency, and level of learning performance). Students in the Hwang et al. study (2015) and Hwang et al. (2011) listened to recordings of peers and pre-recorded how the instructor reads the text in the book. Through sharing, students also exchanged meaningful comments (Hwang et al., 2015). For example, students gave reflective comments and suggestions to a peer who did not complete the homework correctly (Hwang et al., 2011). Comments given by peers were useful in revising and improving homework.
Managing cognitive load in an authentic learning environment

It has been suggested that learning activities which occur in an authentic environment are more likely to facilitate learners’ cognitive activity and conceptual change. The reason is that an authentic environment simulates learners by causing them to interact with real world problems that they are confronted with. This provides learners with real life learning experience and simulates them not only to develop knowledge but also to learn how to apply that knowledge to solve encountered problems. Brown, Collins and Duguid (1989) argue that knowledge is a part of the environment and specific to a particular situation. Moreover, knowledge manifests itself in everyday activities and is in part a product of the activity, context, and culture in which it is developed and used.

From a cognitive load perspective, authentic learning tasks are more complex and impose a high cognitive load on students, especially novices or inexperienced ones. Tasks in a rich real world context have many different solutions, are ecologically valid, and usually cannot be mastered in a single session (Van Merriënboer & Sweller, 2010). Therefore, authentic learning tasks are cognitively demanding and require high-element interactivity, i.e., learners have to deal with several elements simultaneously (Van Merriënboer, Kester, & Paas, 2006). Based on cognitive load theory, instructional design guidelines were proposed in related literature to balance cognitive load and to enhance learning in an authentic environment (Van Merriënboer, 1997; Van Merriënboer et al., 2006; Van Merriënboer & Sweller, 2010). For example, the guidelines suggest to assign learning tasks in simple-to-complex order (i.e., earlier tasks have lower element interactivity), to scaffold learners within a task, and to focus learners’ attention on most important elements for learning. The guidelines also suggest to provide learners with supportive and procedural information (i.e., how a complex problem can be approached and what steps learners need to take) along with part-task practice (i.e., to develop knowledge elements that allow the learner to perform routine aspects at a high level of automaticity). In this way, extraneous load can be decreased, intrinsic load managed and germane load optimized.

In an authentic environment, cognitive tools, such as mobile multimedia applications, can aid learning by diminishing the chances of an excessive cognitive load. Hwang et al. (2013) introduced a mobile learning system to aid sixth grade students’ local culture learning during a field trip taken as part of a social science course. Using the system, students accessed physical and virtual resources in an authentic environment; the system presented the learning tasks, guided students to explore the real-world learning targets, and provided them with supplementary materials via the mobile devices. The effects of this approach on students’ cognitive load and learning achievements were investigated, and the results showed that students who learned with technological approach had better learning achievement and less cognitive load than those who learned using a traditional approach. Hwang et al. (2013) suggest that the mobile learning approach has positive effects on students’ local culture learning.

Elementary school students in Chu’s study (2014) explored an indigenous culture as a part of their social studies course by observing learning objects and completing the learning sheets. Students formed two groups, a control group and an experimental group. The control group participated in learning activities under the instructors’ guidance while the experimental group was guided by a mobile learning system. Chu (2014) investigated the effects of mobile learning on learning achievement and cognitive load and found that learning achievement of the control group was much higher than that of the experimental group. The results also show that there was no difference in extraneous and germane cognitive loads between the two groups; however, the intrinsic load of the experimental group was higher than that of the control group. Chu (2014) concluded that the instructional variables had been well considered in her study, leading to a good balance between controlling the germane and extraneous cognitive loads. However, in an authentic learning environment, students not only need to interact with the real-world learning objects but also pay attention to the guidance and learning content provided by a mobile learning system at the same time. As a result, the learning process becomes more complex and the learning burden (including cognitive load) greater. This is particularly true when learners are under pressure caused by time limitations, the amount of learning content to be covered is large, the number of questions to be answered is considerable, and/or the students’ do not have much experience in a mobile learning environment.

Chen et al. (2008) explored how short-term memory and content representation type affect language learning in a mobile learning environment. To accomplish this, EFL university students were divided into two groups, verbal and visual. They were all required to learn 24 English words (with written annotation and pictorial annotation) delivered by Short Message Service or Multimedia Message Service to their mobile devices. The researchers found that providing vocabulary with pictorial annotation is helpful for learners with lower verbal but higher visual ability as these learners find it easier to learn words presented in a visual rather than in a verbal form. However, providing
vocabulary with both written and pictorial annotation can also help learners with both high verbal and high visual abilities. Chen et al. (2008) concluded that providing the basic learning material is more helpful to learners with low verbal and visual abilities, as too much information may produce a high cognitive load and shorten concentration time.

In a related experiment, Chang et al. (2011) examined how English proficiency (low vs. high) and material presentation mode (single channel vs. dual channel) affect English listening comprehension and cognitive load. In an experimental learning activity, university students studied animals in a zoo using a PDA. The system guided students to target specific animals and then displayed related material (i.e., text) and played an audio guide (spoken messages). Students in a single channel group learned through spoken messages only, whereas students in a dual channel group learned by text and spoken messages. Results of the study revealed that high and low English proficiency learners in the dual channel group had better English listening comprehension than learners in the single channel group. Low English proficiency learners in the dual channel group had a significantly lower extraneous load than those in the single channel group. Chang et al. (2011) concludes that dual channel presentation mode leads to an increased depth of information processing, with the different input modes reinforcing one another.

Research motivation

The literature review of this study evokes four important issues that have not yet been sufficiently addressed. In many countries, English is now the language most widely taught as a foreign language, causing many students to experience cognitive overload. Therefore, how to reduce cognitive overload caused by English learning is an important question for instructors and researchers. Although, many studies have been carried out to address this question, most are limited to vocabulary learning, reading, and listening comprehension; for example, work done by Acha (2009), Chang et al. (2011), Chen et al. (2008), and Plass et al. (2003). Therefore, cognitive load associated with developing such important foreign language skills as speaking and writing has so far been overlooked.

Secondly, the ways in which being in a familiar context (i.e., familiar, relevant and predictable situations from learners’ background and previous experiences) helps comprehension, recall and cognitive load in an authentic environment was not considered in most related studies. Other related studies have created an authentic learning environment by employing mobile technology for research purposes, for example, in a local temple (Hwang et al., 2013; Chu, 2014) or a zoo (Chang et al., 2011). Usually, students visit such places only a few times a year and therefore the learning context there is familiar to students at a certain level. The learning environment used in Chen et al. (2008), though, was only “virtually authentic,” and so was not truly familiar to students.

Thirdly, the research to date has tended to focus on the effects of mobile learning on cognitive load in general rather than how specific mobile technological learning tools can help prevent students from being cognitively overloaded. For example, students unload some important information related to learning objects from everyday life when they use the annotation tool in their tablet PCs. This information would otherwise have had to be constantly held in working memory over a period of time. Later, annotations help students easily recall that important information and build a stronger connection between learning content and targeted objects, i.e., enhanced learning. That is, unloading cognitive work and making more “space” for new information in working memory is essential for keeping the cognitive load balanced, the condition which leads to the most effective learning. This is particularly important for situations in which students’ learning burden can become excessive as they need to interact with real-world learning objects and simultaneously pay attention to learning content provided by mobile technology.

Finally, in previous studies, the ways in which knowledge is explored, verified and applied have received little attention. In most studies, a mobile learning system provides learning content, guides students to the learning targets and displays questions which students are then asked to answer (Hwang et al., 2013; Chang et al., 2011; Chen et al., 2008; Chu, 2014). Such learning processes do not facilitate high level cognitive processing, such as applying new knowledge learned in school to solve problems in daily life situations. Instead they focus on low level cognitive processes, such as recalling and remembering information.

This study strives to overcome these failings in previous research through the design of various learning activities supported by mobile technology. Guided by related literature, learning activities have been designed so students can
learn in class and then freely apply the new knowledge to solve daily life problems in a familiar, authentic environment (e.g., their home or a local convenience store). To make sure that learners are not cognitively overloaded while learning and to enhance their comprehension, this study introduced mobile technology. In this study, non-native English speaking students take pictures of learning objects in a familiar authentic environment and use English to describe them in written and oral annotations by using mobile technology. This study aims to investigate how mobile learning affects learning achievement and cognitive load. The following research questions are addressed in this study.

- Do students who participate in learning activities supported by the system perform better than those without technological support?
- How different is the cognitive load of students during learning activities supported by the system from that of those without technological support?

**Method**

This study employs a quasi-experimental design by adopting a nonequivalent control group method. The effectiveness of applying learning activities supported by the system on learning achievement has been evaluated by comparing the differences in the pre and post-test outcomes of the control and experimental groups in Experiment 1 and 2. The difference in cognitive load on students in the two groups during their learning was also investigated in both experiments. In this study, two classes were selected for the two experiments: one class serves as the control group in Experiment 1 and then becomes the experimental group in Experiment 2; the other class is the exact opposite: it serves as the experimental group in experiment 1 and as the control group in Experiment 2. By assigning two classes, this study hopes to make our results more reliable and valid while investigating the change in learning effectiveness and cognitive load under different conditions.

**Participants and experimental procedures**

Two experiments were carried out in this study. A total of 59 junior high school students participated in the two experiments. One class with 28 students served as the control group in Experiment 1 and as the experimental group in Experiment 2. The other class, with 31 students, served as the experimental group in Experiment 1 and as the control group in Experiment 2. Table 1 presents participants’ profile whereas Figure 1 shows experimental procedure. Most students in both groups were thirteen years old with four to six years’ experience of using computers and less than one to three years’ experience of using tablet PCs. All students had five years of foreign language learning experience prior to this study: three years from the elementary school and two years at the junior high school now. In the primary school, the English education curriculum places greater emphasis on students’ communicative abilities while in the junior high school on developing the reading and writing language skills in addition to communicative abilities (Ministry of Education, 2000).

One may argue that when two classes are recruited for the experiment, half of each class could also be randomly assigned to the experimental and control groups in order to control for potential differences between the classes. In this study, we have assigned one class as one particular group. There are several reasons for this. First, classroom management is an important issue. In junior high school, if some students in class are provided with the technology and some are not then classroom management will be compromised. That is, some students will feel that it is unfair that they have to learn under unequal conditions.

In addition, managing the learning process in a class with many students who use different learning tools will soon become a logistical nightmare for the instructor, who is likely to encounter difficulty controlling and disciplining the class. Second, because students of one class are not acquainted with students of the other, merging them together could hamper their likelihood to take advantage of the sharing functions of the system. That is, students would have no idea whose work is the best to consult in order to learn or to get some inspirational ideas, and they wouldn’t exchange as much feedback as when they learn with peers from the same class due to anxiety to feedback on homework of someone they do not know well.
Learning activity design

This study designed learning activities that focus on learning at school and then applying that knowledge in an authentic environment outside of school, ideally in a wide range of daily life situations (e.g., at local convenience store or supermarket). Learning activities were built around three topics from the textbook: (1) “Where Are You From?” (2) “Your School Is Very Big,” and (3) “Be Quiet and Sit Down, Please” for the first experiment and (1) “Which do you like – Healthy diet,” (2) “How much / many do we need,” and (3) “We were in different classes” for the second experiment. Learning activities include three tasks, and each corresponds to its topic. In each task, all students are asked to take a picture of a learning object (e.g., a sign for Topic 3 of the first experiment or a meal for Topic 1 of the second experiment) and then to introduce and describe it using a minimum of 6-10 sentences.

According to cognitive load theory, learning tasks that carried out in an authentic environment are more complex and impose a high cognitive load on students, especially novices or inexperienced ones. In the study of Chu (2014),
students who learned in an authentic environment using mobile technology experienced low learning achievement but high cognitive load. We learned that the reason for such finding was due to (1) students' simultaneous interaction with the real-world learning objects, the guidance and learning content, (2) high pressure on students caused by time limitations, the large amount of learning content to be covered and the considerable number of questions to be answered, (3) students' insufficient experience in a mobile learning environment. Therefore, in this study, a learning activity supported by mobile technology was devised differently than that in Chu’s (2014) study to ensure better learning achievement and a manageable cognitive load. Furthermore, learning tasks to be carried out in an authentic environment were designed following guidelines proposed by Van Merriënboer (1997), Van Merriënboer et al. (2006), and Van Merriënboer and Sweller (2010). First, students practiced how to apply learned knowledge in the classroom before going to authentic environment. Second, learning tasks were assigned to students in simple-to-complex order. Third, this study ensured beforehand that all the students knew what they needed to do and how to do it during the learning activity and that they had adequate skills to use a tablet PC, that is, that they were able to record their voice and write using the technology. Besides, students learned in familiar to them context and the instructor provided scaffolding and guidance to students when it was necessary. Forth, this study did not create and provide any learning content to students during the learning activity: students were asked to create their own content instead. Fifth, students in this study were asked to complete the required learning activity after school so that there would not be any pressure in terms of time limit or learning content. Finally, our students were not novices or inexperienced ones; instead, they already had five years of foreign language learning experience which included acquisition of communicative abilities as well as reading and writing skills.

The learning system

A learning system was developed for this study which enables students to carry out the learning activity tasks. The interface of the system is shown in Figure 2. The following are the four main functions that the system can execute:

**Annotating.** Students can annotate important parts of the learning material on their tablet PCs. For example, when using the electronic textbook, students can annotate important parts of the learning material, e.g., key concepts, by highlighting them or adding textual explanation to review them afterwards when doing homework or preparing for...
the exam. In addition, students were able to annotate unfamiliar vocabulary in the electronic textbook by adding textual annotations with translation and examples of how to use the new vocabulary in different contexts. When working on tasks, students could write a description of a learning object by creating a textual annotation. Students could also take photos of the learning objects in the familiar authentic environment and attach them to an annotation.

Recording. When students spoke out to describe a learning object, they could record their own voice and listen to it afterwards. In this way, the system enables students to find their own mistakes and improve the content of the recorded audio files. Students could also record the instructor’s lectures and listen to them to review important concepts.

Assistance. Students were able to get assistance from the system, such as (1) reading a text aloud (Text-to-Speech Recognition), (2) translating unfamiliar vocabulary and sentences (Translation), and (3) listing of words in alphabetical order with their meaning and translation (Dictionary).

Sharing. Students were able to share their own annotations, photos, and recorded audio files with peers. This approach allows students to study peers’ annotations so as to both enhance their own understanding of the learning material and improve their own annotations.

In this study, control students completed the same learning activity but by using traditional methods. That is, the control students studied the learning material by using traditional textbooks, took notes in their traditional notebooks and shared notes with others. Control students could also take pictures of learning objects; however, they could not record their voices or the instructor’s lectures.

Instruments

Students’ prior knowledge was evaluated through a pre-test and students’ learning achievement was later measured using a post-test for both experiments. The test items were created by an experienced junior high school teacher based on the learning materials and activities of this study. The items of the pre-test and post-test for both experiments were similar in structure but different in content. Thirty items were included in each test. Test item examples are provided in Appendix 1. Students’ answers to the tests were scored by three raters and notable differences in the assessment were resolved through raters’ discussions until a consensus was achieved. Inter-rater reliability of the tests was evaluated by using Intra-class correlation coefficient (ICC). The average measure ICC was 0.971, indicating high reliability.

Several approaches have been proposed in related literature for measuring cognitive load. Paas, van Merriënboer and Adam (1994) introduced a subjective, indirect measure of cognitive load (a questionnaire survey). In this questionnaire students were asked to report the amount of mental effort they invested in understanding the learning material. Paas et al. (1994) argues that a low amount of invested effort could be a result of low-cognitive load. Kalyuga, Chandler, and Sweller (1999) proposed a subjective, direct measure of cognitive load (a survey that rates the difficulty of the materials). Brunken, Plass and Leutner (2003) recommend an objective, indirect measure of cognitive load based on performance outcomes (knowledge acquisition scores). We adopted a subjective measure of cognitive load for the present study. A cognitive load questionnaire was developed and its design follows the general recommendations from previous related studies (Chu, 2014; Huang et al., 2013; Hwang et al., 2013; Sweller et al., 1998). Four items (shown in Appendix 2) were included in the questionnaire: items 1 and 2 measure cognitive load and items 3 and 4 measure mental effort. All 59 students were asked to respond to the questionnaire, and 59 valid answer sheets were obtained. Responses to the items were scored using a five-point Likert scale, anchored by the end-points “strongly agree” (1) and “strongly disagree” (5). The internal consistency of the survey was tested by employing Cronbach α; the values exceeded 0.80, demonstrating satisfactory reliability of the items. This study has triangulated among different data sources to ground our findings related to cognitive load; that is questionnaire survey results were supported by interview results (Kalyuga et al., 1999) and performance outcomes (Brunken et al., 2003).

One-on-one semi-structured interviews were also conducted with ten students randomly selected from each experiment. Interviews aimed to explore students’ learning experiences with the system and gain insights from their perceptions of the cognitive load they were under. Each interview lasted for 20 minutes and students were asked the
following questions: (1) Please describe your learning experience with the system; (2) Was the system useful for learning? If yes, please explain why. All interviews were audio-recorded after receiving the students' permission, and were then fully transcribed for analysis. The text segments that met the criteria of providing the best research information were highlighted and coded. The codes were then sorted into categories; codes with similar meanings were aggregated together. Established categories produced a framework to report findings pertinent to the research questions.

Results and discussion

The effectiveness of applying learning activities supported by the system on learning achievement and cognitive load was evaluated by comparing the differences in the pre-test, post-test and questionnaire survey outcomes of the control and experimental groups. This study set a priori alpha-level (i.e., level of significance) at 0.05 since an alpha level of less than .05 is accepted in most educational research as statistically significant.

The effectiveness of mobile learning with authentic support on learning achievement

The difference between the pre-test scores of the experimental and control students was measured and then compared by employing an independent samples test. The means and standard deviations of the students' pretest and post-test scores are shown in Table 2. According to the table, students in the control ($M = 47.19$, $SD = 14.71$) and experimental ($M = 48.64$, $SD = 20.81$) groups had equal prior knowledge before the Experiment 1, $t = -0.299$, $p = 0.766$. The table also shows that prior knowledge of the students in the control ($M = 49.03$, $SD = 20.55$) and experimental ($M = 45.46$, $SD = 22.13$) groups was not different before the Experiment 2, $t = 0.637$, $p = 0.527$.

<table>
<thead>
<tr>
<th>Table 2. Results of analysis of covariance with pre-test as covariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
</tr>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>Posttest</td>
</tr>
<tr>
<td>Experiment 2</td>
</tr>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>Posttest</td>
</tr>
</tbody>
</table>

This study used covariance to measure the differences in the learning achievement of students in the control and experimental groups on the post-test, with the pre-test serving as covariate. Results are reported in Table 2. In Experiment 1, a significant difference was observed between the control ($M = 53.50$, $SD = 13.21$) and experimental group ($M = 65.45$, $SD = 18.59$) on the post-test, $F(1,56) = 16.709$, $p = 0.000$, partial eta-squared = 0.236. In Experiment 2, the experimental group ($M = 70.32$, $SD = 17.01$) outperformed the control group ($M = 58.30$, $SD = 22.67$) on the post-test, $F(1, 56) = 20.345$, $p = 0.000$, partial eta-squared = 0.270.

We interviewed a number of students to provide subjective evidence that may support the empirical test results. In the interviews, students indicated that learning activities could be completed more efficiently using the system instead of the traditional approach. Furthermore, the system enabled more effective practice of EFL skills. First, students took pictures of learning objects and created text annotations and recorded their own voice when describing the learning objects. It seems that students enjoyed reviewing the pictures and text annotations and listening to their own recorded files. If the content quality of photos, annotations and recorded files was not satisfactory (e.g., mistakes in pronouncing some words), students wanted to improve it. According to students, such learning behavior led to more frequent language practice as well as to better quality of language output. Similar reasons for using multimedia tools for language practice have been reported in other research (Harmer, 2007; Hwang et al., 2011; Hwang & Shadiev, 2014). For example, students in Hwang et al. (2015) and Hwang et al. (2011) took advantage of the technology to practice the target language repeatedly and regularly. Harmer (2007) found that after students recorded their speeches, they listened to the recordings, evaluated their own language performance, and monitored how much progress they made. However, in contrast to other related research, this study did not only focus on
learning the basic knowledge at school, but also on how students applied the learned knowledge to solve a wide range of real life problems in a familiar authentic environment.

Second, students shared their text annotations and recorded files with peers. In this way, students could read peers’ annotations and listen to peers’ recorded files (i.e., usually to those who study hard and perform well) to get inspirational ideas to complete their own assignments or to learn how peers accomplished assignments so as to improve their own homework. Students could exchange meaningful comments through sharing. That is, some students gave reflective comments and suggestions to a peer who had not completed the homework correctly. Students’ comments were reported to be useful for revising and improving homework. Students thought highly of the sharing mechanism of the system, as they were able to learn from others and use their new knowledge to locate and revise their own homework mistakes. Hwang et al. (2015) and Hwang et al. (2011) argue that, with multimedia aids, students access more diverse learning objects and this may increase the richness of their language experience. They further suggest that sharing multimedia learning content with others not only increases practice opportunities but engages students in EFL contexts and allows them to reflect more deeply on the learning content as well as enter into discussion and collaboration.

Third, students recorded instructor’s lectures. If students forgot some particular parts of a lecture or they needed to listen to the instructor’s pronunciation of the learning material, they could play and listen to the recorded lecture. This was particularly useful outside of the classroom when students could not consult their instructor and ask questions (Hwang et al., 2015).

Finally, students stated that the built-in dictionary was very handy when they needed to translate some unfamiliar vocabulary words when completing assignments outside of school or at home. In this case, a dictionary could help to translate these words. Moreover, with a dictionary, students could find multiple meaning of a word and see how it could be used in different contexts. Hulstijn and Laufer (2001) argue that the use of a dictionary positively affects vocabulary learning. Students look up target words in the dictionary during the reading session in order to find word meanings and to understand the main idea of the target texts. According to Hulstijn and Laufer (2001), students who read foreign language texts and use a dictionary understand the texts better and remember more word meanings.

In this study, students were asked to introduce and describe some learning objects (e.g., signs and rules in the convenience store of their local community) during the learning activity. The students in the experimental group in both experiments completed the assigned tasks better than those in the control group, suggesting that learning activities supported by the system facilitate students learning.

System usage

This section reports how many times experimental students used certain features provided by the learning system in the two experiments. The results are reported in Table 3. According to the table, students in both experiments used text annotation the most. Recorded audio is second, taking photos is third, listening to recorded lecture is fourth and recording lectures is the last in students’ preference list of the system usage. Unfortunately, the system was not designed to record the number of students’ reviews of peers’ text annotations and their listening to peers’ recorded audio. Therefore, the data related to reviewing peers’ text annotations and listening to peers’ recorded audio was not available for analysis. This limitation will be addressed in a future study.

<table>
<thead>
<tr>
<th>#</th>
<th>Text annotation</th>
<th>Students recorded audio</th>
<th>Photo</th>
<th>Dictionary</th>
<th>Listen to recorded lecture</th>
<th>Recorded lecture</th>
<th>Review peers’ text annotations</th>
<th>Listen to peers’ recorded audios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>874</td>
<td>283</td>
<td>198</td>
<td>217</td>
<td>55</td>
<td>69</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>552</td>
<td>259</td>
<td>294</td>
<td>154</td>
<td>89</td>
<td>34</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Total</td>
<td>1426</td>
<td>542</td>
<td>492</td>
<td>371</td>
<td>144</td>
<td>103</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

This study used the data from Table 3 in a stepwise multiple regression analysis to predict post-test scores. This approach provides objective evidence of the system’s functions that are most beneficial for learning achievement.
From Experiment 1 data analysis it was found that at step 1 of the analysis, text annotation was entered into the regression equation, and was significantly related to the post-test scores, \(F(1, 29) = 7.926, p = 0.009\). The multiple correlation coefficient was 0.463, indicating approximately 21.5 percent of the variance of the post-test scores can be accounted for by text annotation. Experiment 2’s data analysis shows similar results; at step 1 of the analysis, text annotation was entered into the regression equation and was found to be significantly related to the post-test scores, \(F(1, 26) = 7.762, p = 0.009\). The multiple correlation coefficient was 0.460, indicating approximately 21.1 percent of the variance of post-test scores can be accounted for by text annotation. Other variables did not enter into the equation at step 2 of either experiments’ data analysis. A hierarchical multiple regression analysis was run to determine the incremental predictive value of text annotation. In this analysis, students’ prior knowledge was added in a first step and the system usage variables in a second. A hierarchical multiple regression analysis demonstrated similar results. That is, text annotation statistically significantly predicted post-test scores in both experiments.

The reason that text annotation was found to be a strong predictive variable may be because students put more effort into creating text annotations and these annotations are easier to modify than other media. For example, the first thing students have to do to complete their tasks is to draft their descriptions of learning objects in text annotations. Students must think thoroughly of how to describe learning objects when drafting their ideas in text annotations, then revise when they get new ideas or inspiration and use these to improve their annotation. Finally, when text annotations are complete, students record their descriptions. Similar finding was reported in other related studies. For example, Hwang et al. (2011) found that only text annotations can significantly predict students learning achievement in Mathematics because text annotations play more important role to learning achievement than any other types of annotations.

### The difference of cognitive load

This study also examined how the designed learning activities supported by the technology system affects cognitive overload during student learning. To accomplish this, the cognitive load of the experimental and control students was measured and then compared by employing an independent samples test. The means and standard deviation from the assessment with respect to seven items on the questionnaire and the results of the \(t\)-test are presented in Table 4. According to the results recorded in the table, the control group students had a higher cognitive load than the experimental group students.

<table>
<thead>
<tr>
<th>#</th>
<th>Control</th>
<th>Experimental</th>
<th>(t)</th>
<th>Sig. 2-tailed</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.05</td>
<td>0.85</td>
<td>1.56</td>
<td>0.44</td>
<td>2.807</td>
</tr>
<tr>
<td>2</td>
<td>2.27</td>
<td>0.83</td>
<td>1.62</td>
<td>0.64</td>
<td>3.325</td>
</tr>
</tbody>
</table>

This study also explores the changes in the cognitive load of the two groups of students when engaged in different pedagogical approaches (learning activities both with and without technological support). For example, group A was the control group in Experiment 1 and used traditional methods to complete the tasks; then, it became the experimental group in Experiment 2 and used technological support to complete the tasks. In contrast, group B was the experimental group in experiment 1 and the control group in Experiment 2. So this study investigated how students’ cognitive load changes when the language engagement method is changed. A dependent samples test was employed for this analysis. The means and standard deviation from the assessment and results of the analysis are presented in Table 5. The results show significant differences in students’ cognitive load when engaged in the two different approaches. Compared to the control students, those who were in the experimental groups had lower cognitive load.

One reason that may explain these two findings is the nature of the learning materials and activities for the two groups. In this study, the learning materials and activities of the control and experimental students were identical apart from that the learning materials for the experimental students were in electronic form, enabling them to take advantage of the system’s functions, (1) Annotating, (2) Recording, (3) Assistance, and (4) Sharing, to complete learning tasks.
Related literature suggests that the intrinsic load lies in the nature of the learning material, learners’ expertise and an interaction between them (Brunken et al., 2003; Sweller et al., 1998). It has been argued that intrinsic load represents the amount of different types of information that students need to consider in the process of acquiring new knowledge, i.e., how much information the working memory needs to deal with at the same time (Mayer & Moreno, 2003). Hwang et al. (2013) argues that intrinsic cognitive load can be affected by the instructional or learning material. He also suggests that students will be cognitively overloaded if the materials are poorly structured, difficult to read, or too complex. In interviews, students in the experimental group mentioned that the system functions seemed to keep the intrinsic load from becoming too heavy when learning with the electronic textbook. Students could annotate important parts of the learning material (e.g., key concepts) by adding textual and multimedia explanation (e.g., a concept meaning and examples of its application in various contexts). Afterwards, these annotations helped students to find important concepts easily, to recall them, and to complete homework or to prepare for exams. It is important to note that the learning material and relevant annotations (i.e., text, photo, and audio) were presented on the same screen. Students anchored their annotations to learning materials which built a connection between them and gave students a clear picture of the whole learning scenario with an appropriate explanation of it. Mayer and Moreno (2003) call this form of presentation an integrated presentation, one which enables learners to focus more on essential information processing. Apart from learning from the electronic textbooks, students can also learn from peers’ annotations. Studying shared annotations, including photos, texts, and audio, helps students to enhance their understanding of the learning material, gives students new ideas and inspiration, and improves their own homework. Students can also get assistance from the system using the Dictionary to find the definition and learn the correct pronunciation of unfamiliar vocabulary. If students need to recall some important concepts taught by the instructor in previous classes, they can listen to recorded lectures.

It has been suggested that extraneous load can be caused by improper instructional design. Thus, in order to reduce extraneous load, instructors need to organize, present and carry out learning information and activities appropriately. In this study, functions of the system helped to reduce the extraneous load when students participated in learning activities. Experimental group students claimed during interviews that when they exposed themselves to the authentic learning environment outside of the school, taking pictures of learning objects and describing them with text or voice annotations, helped them gather their thoughts and then transform them into artefacts, i.e., distributing cognition (Hollan et al., 2000). Lu, Lai, and Law (2010) argue that technology plays an important role in handling intellectual tasks by easing the individual’s cognitive load. Later, when students are at home, they will be in a more tranquil environment in which to study created artefacts (i.e., pictures and their textual and audio descriptions). These artefacts help students to easily recall details of the learning objects from the authentic learning environment, to find out what they missed while completing the tasks, and what else can be improved in their homework. Without this system, students would need to hold a mental representation of the context in working memory over a period of time, which is called “representational holding” (Mayer & Moreno, 2003). In the interviews, experimental group students also claimed that, compared to traditional method, it is easier to participate in learning activities with electronic textbooks and using the system. Students also state that a familiar authentic environment helps them recall vocabulary; the context is related to students’ background and previous experiences and is also highly relevant to learning tasks.

Related literature suggests that the germane load is determined by appropriate instructional design and can enhance learning. With the system’s support, this study has attempted to direct students’ attention to cognitive processes that are directly relevant to the target learning material and tasks. In interviews, experimental group students mentioned that the system functions enabled them to take pictures of learning objects in a familiar authentic environment and then describe them with textual and voice annotations. Students could review their textual descriptions or listen to the audio recorded files afterwards. Students state that completing the tasks in this way facilitated their learning and made it more interesting. Furthermore, students claim that learning in authentic familiar context and creating their own learning materials related to everyday life inspired them to become more engaged with the materials and

<table>
<thead>
<tr>
<th>#</th>
<th>Control M</th>
<th>Control SD</th>
<th>Experimental M</th>
<th>Experimental SD</th>
<th>t</th>
<th>Sig. 2-tailed</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A*</td>
<td>2.02</td>
<td>0.85</td>
<td>1.62</td>
<td>0.64</td>
<td>3.167</td>
<td>0.004</td>
<td>0.402</td>
</tr>
<tr>
<td>Group B**</td>
<td>2.27</td>
<td>0.83</td>
<td>1.57</td>
<td>0.44</td>
<td>5.390</td>
<td>0.000</td>
<td>0.700</td>
</tr>
</tbody>
</table>

Note: *this group was the control group in the Experiment 1 and it was the experimental group in the Experiment 2; **this group was the control group in the Experiment 2 and it was the experimental group in the Experiment 1.
inspired them to try to produce more meaningful output. Huang et al. (2011) recommends increasing students’ interest in learning and engaging them in learning activities and tasks more by utilizing multimedia aids (e.g., pictures and audio). Similarly, Caldwell (1998) argues that multimedia objects in learning stimulate students’ imagination and help them produce meaningful output. Some students, particularly those with lower ability, admit that in the way they learn with the system, they can communicate in the target language with less anxiety about making mistakes. In contrast to traditional learning, students using the experimental system learned with more confidence and their learning was more creative and enjoyable. In interviews, students confirmed that learning content and activities in the electronic textbook with a familiar authentic context were more interesting, fun, and engaging than when using the traditional method.

Based on the above-mentioned results and considering that students in both groups learned with the same learning materials and tasks, this study suggests that learning activities supported by the system enable students to have less cognitive load compared to traditional learning setting.

Conclusion

To address limitations of previous studies in this field, and to facilitate learning and manage cognitive load, learning activities in familiar authentic environment supported by a mobile learning technology were designed, after which two experiments were carried out to investigate how mobile technology support influences learning achievement and cognitive load. During this process, three main findings were made. First, the students using the experimental system outperformed the control students on post-test items in both experiments. Second, learning activities in a familiar authentic environment supported by the system enabled students to have a lower cognitive load compared to the learners without technological support. In addition, according to experimental group students’ reports, multimedia tools helped them to utilize a familiar authentic context better. For example, students could unload some important information from familiar authentic context related to learning objects by using annotating and recording functions of the system. This important information did not necessarily need to be held in working memory over a period of time; it could be easily recalled later and help to build a stronger connection between learning content and targeted objects in familiar authentic context. As a result, the experimental group students’ performance was better and their cognitive load lower. Finally, this study found that creating text annotations is very important learning behavior which predicts learning achievement.

Based on these findings, the authors recommend that educators employ appropriate learning activity designs and use a system which facilitates students’ learning achievement and their cognitive load management. In designing learning activities, the instructor needs to consider how to make the best of the system to develop students’ productive skills and manage their cognitive load. For example, in this study, students took photos of learning objects and described them orally and in writing. Photos and textual and audio descriptions were shared among students so that they could learn from each other and get some new ideas to improve their own homework. The system provided multiple channels for students to present their language output (i.e., taking pictures of learning objects and then describing them in written and oral ways) and gave students more opportunity to use the target language. Thus, instructors can organize learning activities in a way similar to the design used in this study. Furthermore, the instructor needs to encourage students to use the functions of the system, such as annotating, recording, assistance, and sharing, to reduce their cognitive load. In this way, students can efficiently study appropriate learning materials, complete the tasks, and enjoy the learning process at the same time. For example, students can take advantage of annotating to reflect on learning material and review reflections afterwards for better understanding of new concepts or for exam preparation. Students can also distribute their cognition to artifacts created in an authentic environment within a familiar context. Furthermore, the learning environment created by the system can reduce students’ anxiety and help in giving meaningful output, especially for lower ability students. Finally, it is suggested that text annotations are beneficial for learning and therefore students should be encouraged to use them for learning more frequently. For example, if students use text annotations in a familiar authentic context to draft their descriptions of learning objects and later improve them before orally describing them, their performance can be positively affected.

There are several limitations in this study that need to be considered. The first limitation concerns the relatively small sample size. The second limitation relates to short-term exposure of the technology to aid learning. For this reason, these findings cannot be generalized to a broader community based on this study alone or they have limited relevance to learning scenarios in which the technology is used over a longer term when exposed in “real-world”
conditions. These limitations will be addressed in a future study. In the future, our approach can be applied to other domains (e.g., Mathematics or Biology), and cognitive load can be measured objectively by observations of behavior or physiological conditions. A future study will also focus on how a familiar authentic context without mobile learning assistance can decrease cognitive load of students by comparing cognitive load of the control and experimental groups.

Acknowledgements

This research is partially supported by the “International Research-Intensive Center of Excellence Program” of NTNU and the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301, MOST 103-2511-S-006-007-MY3, and MOST 103-2511-S-006-002-MY3.

References


Appendix 1

Test items examples

<table>
<thead>
<tr>
<th>#</th>
<th>Content</th>
<th>Example</th>
<th>Max. score</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 items</td>
<td>Match English word with the correct Chinese meaning.</td>
<td>large 大的&lt;br&gt;fruit 水果&lt;br&gt;pork 猪肉</td>
<td>8</td>
</tr>
<tr>
<td>6 items</td>
<td>Write down the Chinese meaning of English word.</td>
<td>tomato _______________&lt;br&gt;bottle _______________&lt;br&gt;junk food _______________</td>
<td>12</td>
</tr>
<tr>
<td>5 items</td>
<td>Write down:</td>
<td>a) I want two bags of flour. (用 How much 改成问句)&lt;br&gt;b) I’m heavy now. (用 before 代替 now 改写)&lt;br&gt;c) 昨日电影院有很多人，但今天没有.</td>
<td>20</td>
</tr>
<tr>
<td>1 item</td>
<td>Write down.</td>
<td>Write here about yourself when you were at the first grade of the elementary school.&lt;br&gt;Write here about yourself at the moment.&lt;br&gt;Compare and write here the difference between when you were at elementary school and now.</td>
<td>40</td>
</tr>
</tbody>
</table>

Appendix 2

Cognitive load questionnaire

1. Learning these materials was easy.
2. Completing learning activities was easy.
3. Learning these materials did not require a lot of mental effort.
4. Completing learning activities did not require a lot of mental effort.
Effects of Computer-Based Visual Representation on Mathematics Learning and Cognitive Load

Hsin I., Yung and Fred Paas

1Graduate Institute of Technological & Vocational Education, National Taipei University of Technology, Taipei, Taiwan // 2Institute of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands // 3Early Start Research Institute, University of Wollongong, Wollongong, Australia // hcy103@mail.ntut.edu.tw // paas@fsw.eur.nl

*Corresponding author

ABSTRACT

Visual representation has been recognized as a powerful learning tool in many learning domains. Based on the assumption that visual representations can support deeper understanding, we examined the effects of visual representations on learning performance and cognitive load in the domain of mathematics. An experimental condition with visual representations was compared to a control condition without visual representations among primary school students. The hypothesis that learning with visual representations would result in higher learning performance and lower cognitive load than learning without visual representations was confirmed by the results. Theoretical and practical implications of the findings are discussed.

Keywords

Visual representation, Cognitive load, Learning

Introduction

In elementary mathematics, arithmetic operation on integers is generally learned through solving mathematics word problems. It requires students to master number reasoning and number abstraction skills, which are necessary skills for the ability to apply number representation in arithmetical operation on integers (Butterworth, 2006). However, mathematics textbooks do not clearly express how addition and subtraction are related (Van de Walle, 2004; Verschaffel, Greer, & De Corte, 2000). Especially, with regard to arithmetic operation on integers word problems, multiplication and division are highly complex, and without visual representation support, the mathematics concepts tend to be more difficult in addition and subtraction for elementary school children (Múñez, Orrantia, & Rosales, 2013). As a consequence, students are often confused about the relational terms, tend to use all the numbers of word problems, apply the wrong operations principles to reach a final solution, and only possess an instrumental understanding of arithmetic operation on integers principles without relational and logical understanding (Skemp, 1976).

An appropriate instructional approach to promote relational and logical understanding of learning arithmetic operation on integers is using visual representation to emphasize specific relations and promote cognitive functions, such as knowledge construction. Indeed, Barmby, Harries, Higgins, and Suggate (2007) found that the process of understanding can be mediated through visual representations by focusing on relevant information and promoting relational and logical understanding. Visual representation, such as diagrams and pictures, is an externalized form of mathematics concepts and has been recognized as a powerful tool in mathematics learning and problem solving (Ainsworth & VanLabeke, 2004; Gagatsis & Shiakalli, 2004; Zimmermann & Cunningham, 1991). It supports learners’ visualization and interpretation of quantitative data and supports inferences if it contains complementary information that facilitates learners’ cognitive processing (Ainsworth & VanLabeke, 2004). Ainsworth (2006) proposed a conceptual framework and defined three critical functions of multiple external representation including complementary processes, constraining interpretation and construction of deeper understanding. Firstly, the complementary function means that multiple external representations can provide different types of information and facilitate different types of learning processes (Meyer, Shinar, & Leiser, 1997). Different studies on the use of external representations in multimedia learning have shown that they can support learners’ knowledge construction (Ainsworth & VanLabeke, 2004; Ainsworth, 2006). Secondly, the constraining function of multiple external representations holds that familiar representations can be employed to support the interpretation of other representations. Thirdly, multiple external representations can lead to deeper understanding when the information of different representations is integrated. The purpose of using multiple external representations is to provide rich information of a domain and construct references across representations to extend domain knowledge. Based on the above analysis of the functions of multiple external representations and the domain of mathematics, we hypothesized...
that providing learners with visual representations with rich and meaningful information would support them to construct deeper understanding.

Given the importance of the visual representation in learning, considerable research attention has been focused on the role of representations in building abstractions in mathematics understanding and problem-solving ability (Gagatsis & Shiakalli, 2004; Muñez et al., 2013). For example, Gagatsis and Elia (2004) argued that organizational pictures can signal the important information and express the similarities between internal and external forms of the same concepts. In mathematics education, representations are widely used in paper-based form, such as mathematics picture books, in order to develop the story plot, contribute to the text’s coherence and serve as mental scaffolding (Fang, 1996). Levin and Mayer (1993) have proposed four factors that need to be considered when using visual representations in text; the desired performance outcomes, the nature of the representation, the nature of the text and the learner characteristics. Although visual representations can provide rich information and allow learners to go beyond the words, empirical studies have indicated that representations in the form of storybooks may hamper learning, because students have difficulty in extracting and integrating the meaning between words and pictures (Arcavi, 2003; Carney & Levin, 2002). Berends and Van Lieshout (2009) also confirmed that visual representations in a book can negatively impact students’ arithmetic performance, because students not only need to map the elements in the text and representation, but also need to integrate the two sources of information. As a result, additional cognitive load is imposed on learners’ cognitive capacity. Several empirical studies have examined how visual representation in computer-based environments can support students in learning and communicating mathematics concepts efficiently. Ainsworth, Wood, and O’Malley (1998) used a computer based learning environment to teach students multiplication and found that students who learned with representations produced multiple solutions and advanced their understanding.

Learning mathematical operation on integers, such as solving multiplication and division operations word problems, is considered a complex cognitive task, which imposes a high load on students’ working memory, because it involves the representation of the word problem with concrete images, irrelevant information of the problem, selection of an appropriate solution strategy, and application of arithmetic operation principles to arrive at the answer (Dowker, 2005; Muñez et al., 2013; Willis & Fuson, 1988). Zhang and Norman (1994) have shown that visual representations can lead to deeper understanding, because they can relieve some of the limited working memory resources that can be used to establish relationships between abstract mathematic concepts and its underlying functions.

Although several studies related to visual representation have shown that visual representation can improve conceptual understanding in mathematics learning, little is known about the specific effect of visual representation on learning arithmetic operation on integers (Gagatsis & Elia, 2004; Koedinger, Alibali, & Nathan, 2008). Researchers have suggested that using visual representations could enhance learning and reduce working memory load (Elia, Gagatsis, & Demetriou, 2007; Mayer, 2005). Cognitive load theory (Paas, Renkl, & Sweller, 2003; Sweller, Ayres, & Kalyuga, 2011; Sweller, van Merriënboer, & Paas, 1998) provides a solid framework for examining the effectiveness of visual representations and how they can support learners’ cognitive processing (Kalyuga, 2009). Recent studies have shown that visual representation may reduce cognitive load and allow students to redirect the freed resources to processes that can enhance understanding of the problem (Carlson, Chandler, & Sweller, 2003; Kalyuga, 2013).

In light of the analysis above, the present study aimed to examine the effects of the use of visual representation on learning arithmetic operations of integers and the associated cognitive load. Based on cognitive load theory, it was hypothesized that supporting students understanding with visual representation would reduce their cognitive load, enable them to better process the learning materials, and achieve higher learning performance.

**Method**

**Participants**

The participants were 46 fourth-grade students, who were taught by the same mathematics teacher at a public elementary school of Taiwan. All participants were volunteers. Twenty-two of these students were male and twenty-four were female. The mean age of the participants was 10.2 years ($SD = 11.3$ months). Students had been acquainted with basic addition, subtraction, multiplication, and division functions of integer from the third grade.
**Instructional materials**

In this study the instructional materials were based on elementary arithmetic operations of integer including addition, subtraction, multiplication, division, bracket (subtraction formula), bracket (multiplication formula), and two-step addition and subtraction. The goal of the computer-based instruction was to teach the students relational and logical understanding of arithmetic operation on integers and develop their ability to apply the order of operations rules with brackets while solving two-step word problems. The instruction was developed in a story format named “Saving the little Doggy,” which has three main characters and ten scenarios with ten mathematics two-step word problems in a computer-based learning environment. The situated learning scenarios were intertwined with real-life experiences and operation of integer functions, ranging from buying food to dividing goods. The aim of the study was to provide the visual representation of the story-based instruction in a computer-based environment to promote meaningful learning and gradually build relational and logical understanding while solving mathematics problems. As Geary (1995) argued, to build students’ conceptual knowledge of mathematics domain, instruction should involve presenting problems in real world contexts so that students can relate the contents to their own personal experiences. Especially, solving story context problems requires attention for specific information, interpretation and integration of different mathematics operations (Van de Walle, 2004).

Giwawa has twenty cookies and gives Maggie five cookies. Later, Giwawa gives Frankie eight cookies. How many cookies does Giwawa have?

Giwawa algorithm
\[ 20 - (5+8) = ( ) \]

\[ 20 - (5+8) \]
\[ = 20 - 13 \]
\[ = 7 \]

Is the algorithm correct?

*Figure 1a. An example of the story scenario problem for the experimental group*

Giwawa has twenty cookies and gives Maggie five cookies. Later, Giwawa gives Frankie eight cookies. How many cookies does Giwawa have?

Giwawa algorithm
\[ 20 - (5+8) = ( ) \]

\[ 20 - (5 + 8) \]
\[ = 20 - 13 \]
\[ = 7 \]

Is the Giwawa algorithm correct?

*Figure 1b. An example of the story scenario problem for the control group*
An example of the semi-animated story scenario and scenario problem procedures, showing that Giwawa bought cookies and shared his cookies with two friends, for the experimental and control group is presented in Figure 1a and 1b, respectively. Both conditions received the story-based instruction and had the same scenario problems followed by the same problem solving procedures. The only difference between two groups was that the experimental group had the computer-based visual representation and used the arrow to limit learners’ split-attention effect and matching problem between the solving procedures.

**Prior knowledge test**

Numerous studies have shown that learner expertise is a critical factor for instructional design in mathematics (Butterworth, 2006; Kalyuga, 2009). Despite the fact that the participants in this study can be expected to have the same prior knowledge, we used a prior knowledge test and used the score on this test as a covariate in the data analyses to control for possible differences.

A total of 10 computational problems and 10 two-step word problems in paper-and-pencil form were developed for the prior knowledge test. The 10 computational problems involved five multiplication and division operations, and five addition and subtraction operations with brackets. The two-step word problems included mixed multiplication and division functions. To determine the prior knowledge test score, each answer on the prior knowledge test was scored as correct or incorrect based on the final answer. Five points were given for correct answers to each of the 10 computational problems. Five points were given for correct answers to the two-step word problems. If the steps were correct but the final calculation was wrong, three points were given. The maximum score was 100 points. This resulted in the following means and standard deviations for the different conditions: high prior knowledge participants in the experimental group ($M = 87.93, SD = 6.05$), low prior knowledge participants in the experimental group ($M = 67.50, SD = 12.67$), high prior knowledge participants in the control group ($M = 81.67, SD = 6.80$), and low prior knowledge participants the control group ($M = 65.76, SD = 6.80$). The Cronbach’s alpha for the prior knowledge test was 0.76.

**Learning performance test**

The learning performance test was designed to examine students’ comprehension of the arithmetic operations of integer problems. The paper-and-pencil learning performance test consisted of 10 computational problems and 10 two-step word problems. The 10 computational problems involved (1) two-step addition and subtraction with brackets, (2) two-step mixed multiplication and division functions, (3) two-step mixed addition and subtraction and multiplication functions, and (4) questions mixed with a two-step addition and subtraction.

For example, $127 \times (12 \div 3) =$. To determine the learning performance test score, each answer on the learning performance test was scored as correct or incorrect based on the final answer. Five points were given for correct answers to each of the 10 computational problems. Five points were given for correct answers to the two-step word problems. If the steps were correct but the final calculation was wrong, three points were given. The maximum score was 100 points. The Cronbach’s alpha for the learning performance test was 0.75.

**Cognitive load measurement**

Cognitive load was measured using a 7-point scale of perceived difficulty, which is a modified version of the scale developed by Paas and van Merriënboer (1994; see also Paas, Tuovinen, Tabbers, & van Gerven, 2003). Studies have shown that perceived difficulty can be a reliable measure of cognitive load (Kalyuga, Chandler, & Sweller, 1999). Students were asked how difficult the learning material on operations on integers was for them. Cronbach’s alpha for cognitive load was 0.81, which indicated a good internal consistency of the rating scale.
Procedure

The regular mathematics class was taught four times a week, using the national academic edited textbook. The operation of integer unit was taught during the regular scheduled mathematics instructional period for 45 minutes daily. Before the experiment, all students took the prior knowledge test to examine their levels of mathematics knowledge. After taking the prior knowledge test, students were randomly assigned to either the experimental group or control group. All students received the operation of integer function with multimedia story based instruction over five classes. During the instruction phase, the teacher introduced story characters and started with story scenarios. In both conditions students received the same story scenarios. During the problem solving instruction phase, all students were presented with the same 10 arithmetic two-step word problems. Five word problems required multiplication, division with bracket, the other half the word problems required addition, subtraction, multiplication, division with bracket. For example, \(36 \times (9 \div 3)\). The teacher asked the students to read the problems together, and to interpret the meaning of the problem before presenting the next screen. After five minutes, the screen showed how the story characters solved two-step word problems correctly. The only difference between the two groups was that the experimental group received the two-step word problems accompanied by visual representation and the control group received two-step word problems only. After the problem solving instruction, participants were asked to rate the difficulty of the instruction for 1 minute, followed by the paper and pencil learning performance test for 45 minutes.

Results

Learning performance scores and cognitive load ratings were analyzed with one-way Analyses of Covariance (ANCOVA) with instructional condition (with visual representation vs. without visual representation) as between-subjects factor, and prior knowledge serving as a covariate. Table 1 shows the estimated means and standard deviations of learning performance scores and cognitive load ratings for the experimental and control group.

First, homogeneity of regression was determined to test the assumption of the interaction between the prior knowledge and the instructional condition in the prediction of students’ learning performance. The result indicated that the interaction was not significant, \(F(1, 44) = 3.57, p = .06\). After confirming that the data met the ANCOVA assumption, we proceeded with the ANCOVA analysis. The ANCOVA analysis revealed a significant effect for instructional condition, \(F(1, 44) = 5.41, p = .03\), partial eta squared = .28, which indicated that students who learned with visual representations performed better on the test than students who learned without visual representations.

The ANCOVA conducted on the cognitive load ratings revealed a significant effect of instructional condition on cognitive load, \(F(1, 44) = 6.11, p = .02\), partial eta squared = .20 ,which indicated that the students who learned with the visual representations experienced the learning tasks as less difficult than the students who learned without visual representations.

Table 1. Means and standard deviations for learning performance and cognitive load, for students in the experimental and control groups

<table>
<thead>
<tr>
<th></th>
<th>Experimental group (N = 23)</th>
<th>Control group (N = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Learning performance</td>
<td>87.39</td>
<td>8.12</td>
</tr>
<tr>
<td>Perceived difficulty</td>
<td>1.92</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Discussion and conclusion

The present study examined the effects of the use of computer based visual representations in the learning of arithmetic operation on integers on learning performance and cognitive load. The results confirmed the hypotheses by demonstrating that learning with visual representations resulted in higher learning performance and lower cognitive load than learning without visual representations. The results suggest that visual representation, which can convey information about numbers and relations among numbers in a simple form, allowed students to focus attention on the most essential elements. As a result, extraneous cognitive load was reduced, and students could use the freed working memory resources for constructing a coherent mental representation.
Although the perceived difficulty scores that were used as an indicator of cognitive load were differentially affected by the learning conditions, the average scores were very low. This seems to indicate that the students found the task very easy. However, several studies have shown that children generally give low scores, probably because they have the idea that saying that a task was easy, makes you look smarter than saying it was difficult. However, this speculative explanation needs to be verified in future research.

In addition, using visual representations could support learners to process and transform abstract mathematics concepts into concrete representations, and forming such concrete mental images (Sierpinski, 2004). This ability to construct and switch between multiple forms of the same mathematics concepts in the mathematic domain is important because it can support learners to build abstract representation of the concepts and increase the possibility for successful application to new situations, i.e., transfer (Spiro & Jehng, 1990). Therefore, the study confirmed the hypothesis that students learn more with less cognitive load in a learning environment with a visual representation than in a learning environment without a visual representation.

Only a few studies have directly examined the effects of visual representation on learning arithmetic operations of integer from cognitive load perspectives (Berends & Van Lieshout, 2009). We found empirical evidence that using visual representation is beneficial for students in the domain of mathematics. However, more fine-grained studies are needed to examine different types of representations such as dynamic visual representations. By conducting more fine-grained studies, it will be possible to determine the most important characteristics affecting learning and cognitive load. In addition, future research could explore how visual representation facilitates students understanding of abstract mathematics concepts, and how this differs as a function of the level of prior knowledge. It would also be interesting for future research to investigate how level of prior knowledge and working memory capacity mediate the effects of visual representation on learning and cognitive load.

In addition, it would be interesting for future research to disentangle cognitive and affective effects. It is possible that the positive effects of visual representation on mathematics learning were partly caused by the fact that students had more fun and were more motivated to invest mental effort in the instructional materials. In future studies students could be asked to indicate how much they like to work with the instructional materials and whether they would like to work with the same materials in the future.

The relatively small sample size can be considered a limitation of this study. In addition, one specific type of visual representation, one specific topic within a domain, and one specific domain was used in this study. Future studies need to investigate whether other types of visual representations are effective, whether similar effects can be found with other topics within the same domain, and whether similar effects can be found in other domains.

The present study contributes to the knowledge base on the effects of visual representation on learning mathematics. Although the results look promising, it is clear that more research is needed before reliable recommendations can be given for educational practice.

References


Designing Effective Video-Based Modeling Examples Using Gaze and Gesture Cues

Kim Ouwehand1*, Tamara van Gog1 and Fred Paas1,2
1Institute of Psychology, Erasmus University Rotterdam, The Netherlands // 2Early Start Research Institute, University of Wollongong, Australia // ouwehand@fsw.eur.nl // vangog@fsw.eur.nl // paas@fsw.eur.nl
*Corresponding author

ABSTRACT

Research suggests that learners will likely spend a substantial amount of time looking at the model’s face when it is visible in a video-based modeling example. Consequently, in this study we hypothesized that learners might not attend timely to the task areas the model is referring to, unless their attention is guided to such areas by the model’s gaze or gestures. Results showed that the students in all conditions looked more at the female model than at the task area she referred to. However, the data did show a gradual decline in the difference between attention toward the model and the task as a function of cueing: students who observed the model gazing and gesturing at the task, looked the least at the model and the most at the task area she referred to, while those who observed the model looking straight into the camera, looked most at the model and least at the task area she referred to. Students who observed a human model only gazing at the task fell in between. In conclusion, gesture cues in combination with gaze cues effectively help to distribute attention between the model and the task display in our video-based modeling example.

Keywords

Gestures, Video-based human modeling, Eye tracking, Split attention, Cognitive load

Introduction

Over the past decade, learning from videos in which a human model demonstrates and (often) explains how to complete a certain task, has rapidly gained popularity, both in formal and informal educational settings (e.g., YouTube). Such so-called video-based modeling examples provide an opportunity for example-based learning, which is a very effective type of instruction, especially for novice learners (for a review, see van Gog & Rummel, 2010). However, video-modeling examples come in many forms, and little is known about design characteristics that make such examples effective in terms of attention guidance and learning (van Gog & Rummel, 2010). For instance, in video examples in which the model is standing next to a whiteboard or smartboard on which the learning task that the model is explaining is visualized (a typical modern classroom situation), it is possible that the presence of the model creates a type of split-attention effect. The split-attention effect is the adverse effect on learning that is found when students have to mentally integrate information from multiple sources (Ayres & Sweller, 2014). On the other hand, gaze direction and pointing gestures made by the model can automatically trigger attention shifts (Sato, Kochiyama, Uono, & Yoshikawa, 2009). In this way, gaze and gesture cues might be able to timely guide the learners’ attention toward relevant aspects of the learning material and thereby alleviate such split attention. The question addressed in the present study is: What do learners attend to in a modeling example in which the model is visible, and can the model effectively guide learners’ attention by gazing or gesturing at parts of the task?

The model as a potential source of split attention

The reason why seeing the model in the video example might evoke a division of attention between the model and the task that the model is referring to, is that people’s attention is automatically drawn to other people’s faces. There is probably no other object that is looked at as often as the human face, and face perception might well be the most highly developed visual skill in humans, who possess an extensive neural brain circuit involved in face perception and processing (Haxby, Hoffman, & Gobbini, 2000). Moreover, it has been shown that humans prefer to look at faces from a very young age (Tzourio-Mazoyer et al., 2002).

In a study by Gullberg and Holmqvist (2006), in which observers had to listen to and recall an event described by a visible speaker, it was shown that observers focused primarily on the speaker’s face. Eye tracking was used to investigate the amount of viewing time spent looking at a speaker’s face in three conditions: (1) the speaker was telling about the event directly to the addressee, (2) a video (recorded in condition 1) of the speaker was presented at
life-size or, (3) that same video was presented on a 28 inch TV screen. Results showed that over 90% of viewing time was spent looking at the speaker’s face (95.6%, 94.2% and 90.8% in condition 1, 2, and 3 respectively). Although observers had to recall the event the speaker talked about, the speaker did not demonstrate a task, so this study did not investigate how we attend to human modeling examples in which a task is demonstrated and explained to learners.

Even though the findings reviewed above suggest that the model’s face is likely to receive a substantial amount of attention, it is unlikely that learners would look at the model 90% of the time, since they know they have to observe the demonstration and will be tested on their ability to perform that task themselves later on. Indeed, in a recent study using video-based modeling examples in which it was demonstrated how to solve a puzzle problem by manipulating objects (the model was seated behind a table; the puzzle’s objects were placed on the table), half of the participants saw a version of the example in which the face of the model was visible and the other half saw a version of the same example in which the face of the model was not visible. Learners who saw the example video in which the model’s face was visible, were found to look at the model’s face only about 20% of the time, but they outperformed those who did not see the model’s face, after observing the example twice (van Gog, Verveer, & Verveer, 2014). These findings suggest that the attention allocated to the model does not have to result in a negative effect on learning, and that learners are quite able to efficiently divide their attention between the model and the task.

It should be noted though, that in demonstrating this puzzle problem-solving task, the model was gazing at, gesturing at, and manipulating physical objects. This is very different from lecture-style modeling examples in which a model is standing next to a whiteboard on which slides illustrating the steps in the problem-solving procedure are projected and advanced by the model clicking a remote. In such examples, if the model continues to look into the camera, there might be a higher risk of split attention, because learners have to visually search on the screen what the model is talking about, which imposes unnecessary cognitive load during learning (Wouters, Paas, & van Merriënboer, 2008). Furthermore, when learners are looking at the model’s face, they might not attend timely to the task areas the model is referring to, which might result in a) problems integrating the model’s explanation into a coherent mental model of the task, and b) not noticing certain changes in the problem-solving states shown in the slides, especially if the information shown in the slides is transient (i.e., prior steps are no longer visible after each new step/slide is presented; see Sweller, Ayres, & Kalyuga, 2011, on the transient information effect). The question is then, whether we would indeed find evidence that learners may have trouble attending timely to the relevant aspects of the task, and whether gaze cues and gesture cues could help to efficiently guide learners’ attention through such lecture-style video-based modeling examples.

The model’s gaze and gestures as attention guiding cues

In an instructional setting, making deictic gestures (pointing and tracing gestures) has been found to enhance learning (Macken & Ginns, 2014). We suggest that deictic gestures of a video-based model can function as cues to direct learners’ attention toward relevant aspects of the task on crucial moments during the instruction. Research has shown that our attention to faces mainly focuses on the eyes (Vecera & Johnson, 1995) and that eye gaze is a powerful attentional cue; we tend to automatically follow other people’s gaze in order to look at what they are looking at (for reviews see Birmingham & Kingstone, 2009; Langton, Watt, & Bruce, 2000). Indeed, even though the aforementioned study by Gullberg and Holmqvist (2006) showed that in general, speakers’ gestures were hardly fixated at all (less than 1%); observers did relatively often fixate on those gestures that the speakers looked at themselves.

The fact that gestures were hardly fixated in the Gullberg and Holmqvist (2006) study (although it is possible that the gestures were processed through peripheral vision) is quite surprising, because gestures fulfil an important communicative function. For instance, gestures have been found to improve learning (because they capture and guide attention; Valenzeno, Alibali, & Klatzky, 2003) and can communicate information not conveyed in speech (Singer & Goldin-Meadow, 2005). In animations in which a humanoid pedagogical agent gave explanations of the learning content, Mayer and DaPra, (2012) found an embodiment effect, indicating that animated agents producing humanlike behaviour, such as emotional expression, biological movement, gestures and eye gaze, led to better learning outcomes. This effect has also been found with animated pedagogical agents (Moreno, Reislein, & Ozogul, 2010). Moreno et al. (2010) compared learning from a narrated animation with (1) an animated pedagogical agent that produced pointing gestures toward key aspects of the learning material, (2) the same animation in which the gestures...
were replaced with arrow cues, and (3) static visualisations. They found that instruction with a gesturing pedagogical agent led to superior learning compared with instruction using a non-gesturing agent or static visualisations.

Furthermore, research has shown that gestures accompanying speech are perceived as an integrated whole with speech (Kelly, Creigh, & Bartolotti, 2010), and processed in parallel with the head and eye movements (Langton et al., 2000) they accompany. In sum, these results suggest that both gaze and gesture cues are automatically processed and integrated with speech (i.e., quite effortlessly, without imposing much working memory load). These cues might therefore be very useful in video-modeling examples to ameliorate the potential effects of the model’s presence as a source of split attention, by guiding the learners’ attention efficiently through the examples.

The present study

The present study investigated this assumption by measuring learners’ visual attention allocation toward the model and the task aspects in the slides that the model was referring to in her verbal explanation. Participants watched a video-based modeling example showing a human model verbally explaining a novel problem-solving task and either looking straight into the camera (no cue condition), or making occasional gaze shifts toward specific task areas on the screen (gaze cue condition), or making occasional gaze shifts accompanied by pointing gestures toward the screen (gesture + gaze cue condition; see Figure 1 for an impression).

For those scenes of the video-modeling example in which the model was referring to a specific part of the task on the screen (i.e., the small, medium, or large jug), we investigated how learners’ attention allocation (fixation time) was distributed between the model and the task area referred to in that scene. In the scenes in which the model was referring to one of the jugs, students should ideally spend a substantial proportion of time looking at that task Area of Interest (AoI) instead of looking at the model, and cueing might assist them in shifting their focus to the task AoI, with gesture cues being more specific than gaze cues.

It was therefore hypothesized that participants in the no cue condition would spend more time looking at the model and less at the task AoI than those in the gaze cue condition, who would in turn spend more time looking at the model and less at the task AoI than those in the gesture + gaze cue condition. In addition, it was expected that the distribution of attention between the model and the relevant task (screen) areas would be least optimal in the no cue (split-attention) condition, more optimal in the gaze cue condition, and most optimal in the gesture + gaze cue condition. That is, learners in the no cue condition would first have to process what the model was talking about, then shift their attention toward the screen and then search the information on the current slide to determine the right task area, by which time the model might already be at a next step. In the gaze cue condition, distributing attention should be more optimal, because the model’s gaze shift toward the screen would automatically induce an attention shift of the learners, meaning they would look less at the model. However, they might not look more at the task AoI, because they would still have to search for the relevant task area on the current slide as this might not be obvious. This visual search is prevented in the gesture + gaze cue condition, in which attention is not only automatically drawn to the screen, but also to the right aspect of the task on the current slide, which should therefore lead to the most optimal distribution of attention. Besides visual attention, participants’ performance and perceived mental effort...
on subsequent isomorphic and transfer problem solving was measured to explore whether optimal attention
distribution would also lead to optimal performance and effort.

**Method**

**Participants and design**

Participants were 35 Dutch undergraduate Psychology students who participated for course credits. All participants
had normal or corrected-to-normal vision. Despite successful calibration, one participant had to be excluded due to
too much missing eye tracking data, leaving a sample of 34 participants for analysis (20 women, 14 men, $M_{age} = 22.7$
$SD = 1.97$, age range: 20–28).

Participants were randomly assigned to one of three video-based modeling example conditions. In all conditions
participants studied videos of a human model standing next to a screen displaying the problem-solving task, while
verbally explaining and demonstrating the solution procedure that was illustrated by a series of slides projected onto
the screen. Depending on the assigned condition, the model either (1) made no gestures or gaze shifts and looked into
the camera while talking (i.e., no-cue condition), (2), made no gestures, but occasionally looked at relevant task areas
on the screen when these were being mentioned (i.e., gaze cue condition), or (3) looked at and made pointing and
tracing gestures toward the relevant task areas on the screen when these were being mentioned (i.e., gesture + gaze
cue condition). Figure 1 provides an illustration of each instruction condition.

**Materials**

*Problem-solving tasks and video-based modeling examples*

The problem-solving task consisted of an adapted version of the water-redistribution paradigm of Schmid, Wirth, and
Polkeln (2003), which is based on Luchins’ (1942) water jug task. Participants were presented with three jugs with a
certain maximum content (displayed above each jug) containing certain amounts of water (displayed inside each
jug), which they were instructed to redistribute until a goal state (displayed below each jug) would be reached (see
Appendix 1 for an example). Problem solving was constrained by one task rule: The entire content of the donating
jug would always be emptied into the receiving jug (i.e., no partial contents could be redistributed), unless the
receiving jug would not have enough capacity for the content of the donating jug, in which case the receiving jug
would be filled to the brim, leaving the donating jug with the residual. The problems used for the present experiment
consisted of three-step water-redistribution problems that could only be solved with a counterintuitive strategy.
Carder, Handley, and Perfect (2008) explain the counterintuitive strategy with the evaluation factor (EVF), which is
the sum of differences between the current and goal states of all jugs. For example, in Figure 1, the EVF is 6 ($3 + 1 +$
2). A step that decreases the EVF is called perceptually consistent, because it directly brings the problem solver
perceptually closer to the goal state. A counterintuitive step increases the EVF, but is sometimes a necessary step in
the solution pathway. Hence, problems that should be solved with a counterintuitive strategy require problem solvers
to look more than one move ahead (Bull, Espy, & Senn, 2004) and are therefore more demanding for working
memory than problems that can be solved with a perceptually consistent strategy (Carder et al., 2008).

A computerized version of this water-redistribution task (Schmid et al., 2003) was created in E-prime 2.0.
Participants could redistribute water through mouse clicks on the jugs. In Figure 1, for example, in order to pour
water from jug A into jug B, participants first had to click on the jug they wanted to pour water from (i.e., the
donating jug, in this case A; the water in this jug changed to a darker color as a visual confirmation that it was
selected) and secondly, on the jug they wanted to pour water into (i.e., the receiving jug, in this case B). With the
second click, the water levels of the jugs changed according to the task rule.

For each instruction condition, a video-based modeling example was created, in which the same female model
explained a problem-solving task while standing next to a screen depicting the task (a typical lecture situation). In all
three conditions, the model gave the same verbal explanation (see Appendix 1). The problem state depicted on the
slide that was projected on the screen changed automatically to the next problem state (i.e., slide) when the model
mentioned a problem-solving step being performed, so no interaction of the model with the screen was required. The
video examples in all conditions were divided in 33 scenes, consisting of six scenes in which participants were expected to look at the model (because no task-relevant areas on screen were referred to), and 27 scenes in which the model referred to task-relevant areas. Task-relevant areas were referred to verbally in the no cue condition, verbally combined with simultaneous gaze shifts in the gaze cue condition, or verbally combined with simultaneous gaze shifts and gestures in the gesture + gaze cue condition (see Figure 1). The video-based modeling examples were recorded with a digital video camera and edited in Final Cut Pro 7.0.3. All videos had the same duration of two min and were presented in E-prime 2.0.

Mental effort

After each problem participants rated how much mental effort they invested in solving it, which is an indicator of experienced cognitive load. The mental effort rating scale consisted of labeled values ranging from 0 (no effort) to 9 (extremely high effort) and was adapted from Paas (1992; see also Paas, Tuovinen, Tabbers, & van Gerven, 2003). The mental effort rating scale was also presented in E-prime 2.0 and participants responded by pressing a number on the keyboard that corresponded to the amount of mental effort they perceived to have invested in the task.

Eye-tracking equipment

The video-based modeling examples and problem-solving tasks were presented in E-prime on the 21-inch display of a Tobii 2150 (50 Hz) eye tracker, which registered participants’ eye movements while they studied the modeling examples. Participants sat approximately at a 70 cm distance from the screen. To show the videos full screen they were presented with a 600 x 800 resolution. The system was recalibrated in IView prior to each example, with a 5-point calibration.

Procedure

The experiment was conducted in individual sessions of approximately 15 minutes. Participants first read a short written instruction about the basic task rules, for which they received three min. Subsequently, the system was calibrated and participants were instructed to sit as still as possible while they studied the modeling example for the first time. They were then presented with an isomorphic problem to solve (during which they could move freely). They were then presented with a new isomorphic problem to solve (during which they could move freely) after which they rated how much mental effort they perceived to have invested in solving that problem. Finally, participants were presented with two transfer problems, in which the same procedure could be used to solve the problem, but the jugs had different positions, so participants could not just copy the procedure exactly as they observed it. Each transfer problem was followed by the mental effort rating scale. Participants received a maximum of one min per problem for all problems presented during the experiment.

Data analysis

Eye-movement data

The video examples were divided into 33 scenes. There were two types of scenes. In six scenes the main Area of Interest (AoI) was the model (if she was providing explanations not directly referring to the task) and in 27 scenes the main AoI was a part of the task, that is one of the three jugs (with the accompanying numbers above, in and under the jug) that the model was referring to either verbally only, verbally with gaze shifts or verbally with gaze shifts and gestures (see Figure 1). Fixations were defined as gaze points that fell within a radius of 30 pixels and together had a duration of more than 60 ms, and for each AoI in each scene, fixation duration was calculated. Fixation duration on the model and fixation duration on the relevant task area in each scene (i.e., the area being referred to by the model in that scene, which could vary across scenes) were summed only for those 27 scenes in which the task area was the main AoI, and subsequently transformed into a percentage of total fixation duration on those scenes.
Learning outcomes

For each isomorphic problem solved, a performance score was computed by dividing the number of steps in the shortest possible solution (i.e., three), by the actual number of steps a participant took to solve the problem. For example, if participant A solved a three-step problem in three steps and participant B solved the same problem in 15 steps, this would result in a score of 1 for A and 0.2 for B. The same formula was applied for transfer performance, but here, one score was obtained by determining the average performance score over the two transfer problems.

Results

Eye movement data

Table 1 shows the means and standard deviations for the fixation duration (percentage) on the model, on the task areas she referred to (averaged across scenes; hereafter called relevant task area), and on the remaining task areas (averaged across scenes). Because the overall data include those scenes in which the model did not refer to aspects of the task, the data from the task scenes only are more relevant for our hypothesis, and these were analysed with a 3 x 2 x 2 ANOVA with instruction condition (no cue, gaze cue, and gesture + gaze cue) as between-subjects factor and object of attention (model vs. task) and time (first example vs. second example) as within-subjects factors.

Table 1. Percentage fixation duration in task scenes

<table>
<thead>
<tr>
<th>Time</th>
<th>Object of Attention</th>
<th>No cue ( (n = 11) )</th>
<th>Gaze cue ( (n = 12) )</th>
<th>Gesture + Gaze cue ( (n = 11) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td>First</td>
<td>Model</td>
<td>45.86</td>
<td>17.59</td>
<td>33.72</td>
</tr>
<tr>
<td></td>
<td>Relevant Task Area</td>
<td>12.81</td>
<td>6.18</td>
<td>17.94</td>
</tr>
<tr>
<td></td>
<td>Remaining Task Areas</td>
<td>24.76</td>
<td>8.47</td>
<td>25.23</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>16.57</td>
<td>9.05</td>
<td>23.10</td>
</tr>
<tr>
<td>Second</td>
<td>Model</td>
<td>40.75</td>
<td>23.15</td>
<td>35.17</td>
</tr>
<tr>
<td></td>
<td>Relevant Task Area</td>
<td>15.01</td>
<td>10.56</td>
<td>16.77</td>
</tr>
<tr>
<td></td>
<td>Remaining Task Areas</td>
<td>24.77</td>
<td>11.04</td>
<td>24.99</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>19.47</td>
<td>10.01</td>
<td>23.07</td>
</tr>
</tbody>
</table>

Note. Fixations on “other” areas are fixations to white space above, below, or next to the task and the model.

Figure 2. Interaction between instruction condition and object of attention (error bars represent standard errors + 2 SE)

The analysis showed no main effect of instruction condition, \( F(2, 31) = 1.51, MSE = 184.07, p = .236, \eta^2 = .09 \), or time, \( F(1, 31) = 0.24, MSE = 33.55, p = .629, \eta^2 < .01 \), a main effect of object of attention, \( F(1, 31) = 39.08, MSE \)
= 299.28, \(p < .001\), \(\eta^2_p = .56\), and an interaction between object of attention and instruction condition, \(F(2, 31) = 3.81, MSE = 299.28, p = .033, \eta^2_p = .20\) (see Figure 2). There was no interaction of instruction condition and time, \(F(2, 31) = 0.24, MSE = 33.55, p = .786, \eta^2_p = .02\), object of attention and time, \(F(2, 31) = 0.63, MSE = 129.85, p = .560, \eta^2_p = .04\).

We followed up on the significant Instruction Condition x Object of Attention interaction with multiple comparisons between instruction conditions on the attention distribution between model and task-relevant areas. We calculated a measure of attention distribution by subtracting the total fixation duration toward task-relevant areas from the total fixation duration toward the model. Results show a significant difference between the no cue and the gesture + gaze cue group, \(t(20) = 2.57, p = .023, d = 1.10\), but no difference between the no cue and gaze cue group, \(t(21) = 1.48, p = .153, d = 0.61\), or the gaze cue and gesture + gaze cue group, \(t(21) = 1.47, p = .157, d = 0.62\). These results indicate that participants in the gesture + gaze cue group had a smaller attentional bias toward the model compared with the task-relevant areas than participants in the no cue group. Figure 2 depicts the interaction between instruction condition and object of attention.

Learning outcomes

Table 2 shows the means and standard deviations of the performance and mental effort data on the isomorphic and transfer problems. Performance and mental effort measures of isomorphic problem solving were analyzed by 3 x 2 mixed ANOVAs with instruction condition (no cue, gaze cue, and gesture + gaze cue) as between-subjects factor and time (problem solving after the first and second time participants watched the video) as within-subjects factor. Performance and mental effort measures of transfer problem solving were analyzed by ANOVAs with instruction condition (no cue, gaze cue, and gesture + gaze cue) as between-subjects factor.

Isomorphic problem solving performance and mental effort

For performance, results showed no main effect of instruction condition, \(F(2, 31) = 2.86, MSE = 0.23, p = .072, \eta^2_p = .16\), a main effect of time, \(F(2, 31) = 26.21, MSE = 0.13, p < .001, \eta^2_p = .46\), but no interaction, \(F(2, 31) = 0.53, MSE = 0.13, p = .595, \eta^2_p = .03\). The analysis of perceived mental effort invested in solving the isomorphic problems showed no main effect of instruction condition, \(F(2, 31) = 1.66, MSE = 12.25, p = .208, \eta^2_p = .10\), a main effect of time, \(F(1, 31) = 13.80, MSE = 6.08, p = .001, \eta^2_p = .31\), but no interaction, \(F(2, 31) = 0.45, MSE = 6.08, p = .643, \eta^2_p = .03\). As Table 2 shows, these results reflect improved performance and decreased perceived mental effort on problem solving after the second compared with the first example.

Transfer problem solving performance and mental effort

Results showed no main effect of instruction condition on transfer performance, \(F(2, 31) = 1.35, MSE = 0.20, p = .275, \eta^2_p = .08\), or perceived mental effort invested in solving the transfer problems, \(F(2, 31) = 0.54, MSE = 9.33, p = .588, \eta^2_p = .03\).

Table 2. Learning and transfer performance, and effort

<table>
<thead>
<tr>
<th>Condition</th>
<th>No cue (n = 11)</th>
<th>Gaze cue (n = 12)</th>
<th>Gesture + Gaze cue (n = 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>0.42</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>Second</td>
<td>0.81</td>
<td>0.39</td>
<td>0.90</td>
</tr>
<tr>
<td>Transfer</td>
<td>0.69</td>
<td>0.46</td>
<td>0.65</td>
</tr>
<tr>
<td>Mental Effort</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>5.33</td>
<td>3.39</td>
<td>4.50</td>
</tr>
<tr>
<td>Second</td>
<td>3.17</td>
<td>2.69</td>
<td>2.83</td>
</tr>
<tr>
<td>Transfer</td>
<td>3.75</td>
<td>3.00</td>
<td>3.54</td>
</tr>
</tbody>
</table>
Conclusion

The present study focused on the question of whether gaze and gesture cues would improve the distribution of visual attention when studying a video-based modeling example in which a human model explained how to solve a novel problem. The data showed a clear trend in line with our hypothesis that students looked more at the model than at the task-relevant AoI, and that gaze and gesture cues can help shift attention from the model to what she is talking about; students in the no cue condition, looked most at the model and least at the task, while students in the gesture + gaze cue condition looked most at the task and least at the model compared with the other two conditions, and the gaze cue condition falling in between. Thus the attention toward the model gradually decreased and the attention toward the task gradually increased from the no cue, to the gaze cue to the gesture + gaze cue condition. Or in other words, participants who learned from a human model that occasionally gestured and gazed toward the task screen had a smaller attentional bias toward the model compared with the task-relevant areas than participants that learned from a model that did not gesture or gaze at the task.

Our main focus in this study was on the effect of gaze and gesture cues on learners’ visual attention distribution, but we also explored whether type of instruction affected learning outcomes, although, in contrast to the eye movement data, for the performance and mental effort data this sample size was probably too low to have sufficient power to detect possible differences. Indeed, we found no significant effects on learning outcomes as measured by performance and mental effort on the isomorphic and transfer problems. It is a likely assumption that students who spent more time looking at the model than at what the model is talking about would not be able to smoothly integrate the visual and verbal information provided in the example and that this would hamper their learning (see also Mayer & DaPra, 2012; Moreno et al., 2010). Therefore, future research should replicate this experiment with larger sample sizes in order to address the question of whether better distribution of visual attention between the model and the task-related areas she is referring to, would indeed improve learning.

In sum, this study confirmed that when learning from videos, the model’s face attracts a substantial amount of learners’ attention, and showed that providing cues, gestures in particular seem effective in redirecting learners’ attention from the model to the task areas the model is referring to. Given that the use of lecture-style online instructional videos is rapidly increasing, these findings contribute toward the development of design guidelines for such videos.

References


Appendix 1

Verbal script of the video

“The next problem can be solved in three steps. The correct solution can be found if you focus on the goal amount of the large jug. You can see the solution in the next formula: the current quantity of the large jug – the quantity that can be added to the medium jug + the maximum quantity of the small jug, or \(7 - 2 + 4 = 9\).

The first step is to pour water from the large jug into the medium jug. The medium jug will reach a quantity of \(4 + 2 = 6\). The large jug will reach a quantity of \(7 - 2 = 5\).

After the first step, the jugs look like this.”

Next slide appears.

“The second step is to pour water from the medium jug into the small jug. The medium jug will reach a quantity of \(6 - 1 = 5\), which is equal to its goal amount. The small jug will reach a quantity of \(3 + 1 = 4\). After the second step, the jugs look like this.”

Next slide appears.
“The final step is to pour water from the small jug to the large jug. The small jug will reach a quantity of $4 - 4 = 0$, which is equal to its goal amount. The large jug will reach a quantity of $5 + 4 = 9$, which is equal to its goal amount. After the final step, the jugs look like this.”

Next slide appears

“*The problem is now solved.*”

End of video.
Computer-Based Learning of Geometry from Integrated and Split-Attention Worked Examples: The Power of Self-Management

Sharon Tindall-Ford1, Shirley Agostinho2*, Sahar Bokosmaty2, Fred Paas2,3, and Paul Chandler2

1School of Education, University of Wollongong, Australia // 2Early Start Research Institute, University of Wollongong, Australia // 3Erasmus University, Rotterdam, The Netherlands // sharontf@uow.edu.au // shirleya@uow.edu.au // saharb@uow.edu.au // paas@fsw.eur.nl // chandler@uow.edu.au

*Corresponding author

ABSTRACT

This research investigated the viability of learning by self-managing split-attention worked examples as an alternative to learning by studying instructor-managed integrated worked examples. Secondary school students learning properties of angles on parallel lines were taught to integrate spatially separated text and diagrammatic information by using online tools to physically move text to associated parts of a diagram. The moving of text aimed to reduce learners’ need to search between text and diagram, freeing cognitive resources for learning and affording learners’ control of their learning materials. The main hypotheses that learners who self-manage split-attention worked examples would perform better on test items than learners who study split-attention worked examples, and perform as well as learners who study instructor-managed integrated worked examples were confirmed. Theoretical and practical implications of the results are discussed.

Keywords

Cognitive load theory, Split-attention, Self-management of split-attention, Worked examples

Introduction

Today’s online learning environment provides learners with access to information of variable quality, both academically and structurally. Learners need to be able to assess the academic quality of information, recognize poorly constructed materials, and reorganize information to support their learning. This paper explores the efficacy of teaching learners to reorganize information by manually integrating related text within a diagram to reduce search and support their understanding of mathematical concepts, thus decreasing the load on working memory and enhancing learning. The research was informed by Cognitive Load Theory (CLT) (Paas, Renkl, & Sweller, 2003; Sweller, 1988; Sweller, Ayres, & Kalyuga, 2011), which posits that effective instructional design should take into account human cognitive architecture, specifically focusing on effective use of limited working memory resources (Ayres & Paas, 2008). CLT instructional designed materials aim to remove processing of superfluous information (extraneous cognitive load), manage the difficulty of the information to be learnt (intrinsic cognitive load), while optimizing the learners’ processing capacity for what is relevant to learning (germane cognitive load) (Sweller, 2010). Traditionally, CLT research has focused on the instructor/teacher-designed learning materials that fulfill the above. Some of the best-studied instructional approaches include the worked example effect (Atkinson, Derry, Renkl, & Wortham, 2000; Sweller, Ayres, & Kalyuga, 2011). This effect occurs when learning is enhanced by studying worked examples which are stepped-out solutions rather than being presented with the equivalent problem with no stepped-out solution. Another instructional principle is the split-attention effect, which was found when studying a specific type of worked example, consisting of mutually dependent but spatially separated text and diagram, and solving the equivalent problems were equally ineffective for learning (Sweller, Chandler, Tierney, & Cooper, 1990). The split-attention effect is defined as the lowering in learning performance caused by presenting mutually dependent, but spatially separated sources of information (e.g., text and diagram) in the visual modality, requiring the learner to invest working memory resources to integrate the dispersed information. This study investigated a new perspective on the split-attention effect, in which learners are taught to self-manage sources of split-attention in worked examples.

The split-attention effect occurs when a learner is forced to search and mentally integrate related information that is incomprehensible in isolation due the design of the instructions (Ayres & Sweller 2005). It is the searching and mental integration undertaken by the learner that places an extraneous load on limited working memory resources and inhibits learning (Chandler & Sweller, 1991). Aligned with the split-attention effect are the spatial and temporal contiguity effects (Mayer, 1997; Mayer & Moreno, 2003; Moreno & Mayer, 1999). Both effects focus on the importance of ensuring that related information be close in proximity in regard to space (spatial contiguity effect) or...
time (temporal contiguity effect). Ginns’ (2006) meta-analysis on the importance of reducing learners’ split-attention showed that integrated instructions led to more efficient and effective learning compared to instructions presented in a split-attention format. An example of split-source materials may include a diagram with related written text presented above, below or to the side. Under split source instructional conditions the learner is forced to mentally hold and integrate in working memory the related diagram and text in order to process and understand the instructions (see Figure 1). It is the division of cognitive resources that poses an extra load on limited working memory, which impedes learning. This can be mitigated by integrating related information (see Figure 2). It is argued that, as learners are no longer required to mentally search and match text with diagram, extraneous load is reduced and cognitive resources are freed for learning. The benefits of integrated instructions as an alternative to split source materials has been demonstrated across a range of learning domains including mathematics (e.g., Tarmizi & Sweller, 1988), physics (e.g., Ward & Sweller, 1990), engineering (e.g., Mayer, 1989) and second language learning (e.g., Chung, 2008). Traditionally the research has relied on an expert instructor designing the integrated instructions. The novelty of the current research is the investigation of the efficacy of teaching learners to meaningfully integrate text with a diagram to reduce search and support learning.

Self-management as an alternative to instructor management of cognitive load is a new development in CLT research (Agostinho, Tindall-Ford, & Bokosmaty, 2014; Agostinho, Tindall-Ford, & Roodenrys, 2013; Roodenrys, Agostinho, Roodenrys, & Chandler, 2012). Self-management in this current experiment required the learner to move textual information to parts of the diagram which the learner understands are related, the aim being to reduce learners’ need to search and freeing cognitive resources for learning. Despite learners in the self-managed condition being required to undertake an additional task to integrate text and diagram, this task may serve to not only reduce searching between related sources of information but also focus learners attention on the to be learnt material thus increasing germane load (see also Paas & Van Gog, 2006; Paas & Van Merriënboer, 1994). Roodenrys et al. (2012) investigated the effectiveness of teaching self-management skills to tertiary students. Two experiments in the domain of educational psychology showed the superiority of integrated instructions compared to traditional split-attention materials. The research provided some evidence to suggest that self-management may be an alternative to instructor managed integrated instruction. An interesting finding was that teaching the strategy of self-management was transferable, with the self-management group significantly outperforming both integrated and split-attention groups on test items in a different learning domain. Follow-up research on self-management in an online environment with tertiary students was undertaken by Agostinho et al. (2013). The research, conducted with pre-service teachers in an educational technology subject, utilized both qualitative and quantitative measures to investigate the efficacy of self-managing split-attention online. Quantitative results demonstrated that integrated instructions were superior to split source and self-management instructions, confirming a split-attention effect. Results showed that although there was not a significant difference in favor of the self-management group, this group did perform better on all test items compared to the split-attention group. This result was despite the fact that the self-management group was required to undertake the additional task of moving text to integrate with the diagram during the learning phase. Qualitative results, which were obtained by verbal protocols, suggested that explicit instruction and training are required for learners to understand how to successfully integrate text with diagram to self-manage split-attention. Thus, the preliminary research conducted to teach students to manage split-attention themselves indicates that learners who are guided on how to physically manipulate either print or computer based instructional materials with evident split-attention perform as well, and on some test items better, than learners learning from split-source instructions.

Another potential advantage for enabling learners to self-manage split-attention materials is that it provides a form of learner control which can lead to a higher level of learner engagement (Orvis, Fisher, & Wasserman, 2009). This is a form of germane cognitive load, which potentially enhances learner understanding (Hasler, Kersten, & Sweller, 2007). It is speculated that self-management requires learners to actively engage with the “to be learned content,” which may lead to enhanced processing and schema construction. However a meta-analysis on the positive benefits of learner control in computer-based training indicated the benefits to learning are relatively small (Kraiger & Jerden, 2007). Granger and Levine (2010) suggested for novice learners when first learning complex information in a computer based learning environment, it is important that the learner has only minimal control and decision-making. Research indicates that when information to be learnt is high in complexity (high element interactivity) with a corresponding high working memory load it is essential not to further burden novice learners’ working memory with additional activities or decision making that are not aligned with learning (e.g., Bokosmaty, Sweller, & Kalyuga, 2014). To optimize self-management of the split-attention effect in this research, novice learners were given limited choices and decisions to make when manually integrating related text within a diagram. Furthermore, specific
guidance and training on the reasons why and how to self-manage the split-attention effect was integral to the experimental design.

The current study investigated the effects on secondary students’ geometry learning of a self-management strategy for cognitive load by physically manipulating digital instructional materials with evident split-attention. Similar to the previous research on self-management of cognitive load (Agostinho et al., 2013; Roodenrys et al., 2012), the effectiveness was determined by comparing the self-management condition to an instructor-managed condition with physically integrated instructional materials and no split-attention, and to a traditional split-attention condition.

Firstly, it was hypothesized that participants studying split-attention worked examples, who are provided with guidance on how to self-manage split-attention, would outperform participants studying split-attention worked examples on a transfer test (Hypothesis 1). Secondly, it was hypothesized that participants studying instructor-managed integrated worked examples (that is, with no evident split-attention) would outperform participants studying split-attention worked examples on a transfer test, thereby replicating the split-attention effect (Hypothesis 2). Thirdly, it was hypothesized that participants studying split-attention worked examples and provided guidance on how to self-manage split-attention would perform equivalent to participants studying instructor-managed integrated worked examples on a transfer test (Hypothesis 3). Fourthly, with regard to cognitive load during learning, it was hypothesized that participants studying split-attention worked examples and participants self-managing split-attention worked examples would have higher cognitive load ratings for studying instructions than participants studying instructor integrated worked examples (Hypothesis 4), due to higher extraneous load in the split-attention condition and higher germane load in the self-management condition. Fifthly, with regard to cognitive load during the transfer test, it was hypothesized that participants in the split-attention condition would report higher cognitive load than participants in the instructor-managed and self-managed conditions (Hypothesis 5), because of the higher quality of schemata constructed in the latter two conditions during learning.

Method

Participants and design

Participants were 48 female Year 7 students (between 12 and 13 years of age) from a New South Wales secondary school. The students and their parents/carers consented to participate in the study. All students had participated in an introductory lesson (designed and delivered by one of the researchers) to revise concepts such as parallel lines and introduce the concept of alternate and co-interior angles on parallel lines. This lesson however, did not explain how to calculate corresponding and alternate angles on parallel lines, as this was the focus for the experiment. The experiment was conducted across eight school days over a period of two school weeks during the third of four 10-week school terms. Participants were randomly assigned to one of three conditions:

- Split-attention
- Integrated
- Self-managed

Materials and procedure

The instructional material was based on New South Wales Board of Studies Mathematics K-10 Stage 4 (Years 7 and 8) syllabus outcomes (Stage 4, Measurement and Geometry: Angle Relationships in Board of Studies, 2012). The topic of properties of angles on parallel lines was selected because students were scheduled to study the subject matter in the subsequent school term, thus the experiment served as an introduction to these mathematical concepts.

Prior to the commencement of the experiment all students in the four mathematics classes received a 45-minute lesson taught by one of the researchers. In this lesson the researcher revised geometry concepts such as parallel lines, transversal lines and angles on straight lines adding up to 180°. These concepts had been introduced to the students earlier in the year (in the first school term) and thus, this part of the lesson served as revision. The second part of the lesson introduced the two new topics: alternate and co-interior angles on parallel lines. Participants completed a worksheet during the lesson to reinforce the content presented in the lesson. Students participated in the experiment one week after participating in this introductory lesson. The experiment took fifty minutes and was undertaken in
groups of two to three participants from the same instructional condition with one researcher in a large tutorial room. Each participant worked independently at a designated laptop computer and was unable to view other participant’s computers or work. The instructional materials were developed using an interactive whiteboard software application called SMART Notebook. This software application (similar to Powerpoint) enables text and shapes to be moved on the screen as objects. Three SMART Notebook files were developed (one for each condition) each consisting of 10 slides. The instructional materials were formatted as follows.

**Split-attention:** Instructional materials comprised of worked examples presenting parallel lines with a number of angles indicated on the top half of the screen followed underneath by a series of textual statements explaining the relationship between and the calculation of the angles (see Figure 1). The text and diagram were “locked” and were not movable.

**Integrated:** Instructional materials included worked examples showing the same parallel lines, angle and textual statements, as in split-attention condition, however the textual information was integrated on the diagram to reduce split-attention (see Figure 2).

**Self-managed:** Instructional materials included worked examples showing the same parallel lines, angle and textual statements, as in split-attention condition, however the text statements were movable (see Figure 3).

---

**Task 1:** Find angle $a$

Read the text and look at the diagram to see how the text helps you understand how to find angle $a$.

$\text{Figure 1. Example of a split-attention worked example (split-attention condition)}$

---

**Task 1:** Find angle $a$

Read the text to see how it relates to that part of the diagram so you understand how to find angle $a$.

$\text{Figure 2. Example of an integrated worked example (integrated condition)}$
The experiment comprised four phases, which are described below.

**Phase 1: Revision of concepts taught in lesson (5 minutes)**

Participants worked through four SMART Notebook slides to revise the concepts taught in the previous week’s mathematics lesson on parallel lines and angles. All participants, irrespective of their allocated condition, worked through the same four SMART Notebook slides. The first SMART Notebook slide provided information on parallel and transversal lines, the second slide showed how two angles on a straight line add up to $180^\circ$, the third slide showed the different positions of alternate angles on parallel lines and how alternate angles are equal. The final slide showed the different positions of co-interior angles on parallel lines and how they add up to $180^\circ$.

**Phase 2: Training of instructional condition: (10 minutes)**

The training phase comprised two Notebook files where participants were presented with two worked examples based on the instructional condition to which they were allocated. The worked examples were focused on how to calculate angles on a straight line when presented with 2 parallel lines and 1 transversal line. Each worked example provided a two-step solution on how to calculate an angle on a straight line. The two-step solution was presented as 2 text statements. For the split-attention and self-managed conditions the two text statements were presented underneath the diagram. In the integrated condition the two text statements were positioned in the relevant parts of the diagram. Participants were provided with verbal instructions from the researcher on how they were to study the worked example based on the instructional condition to which they were allocated. Based on the results from the pilot study, the verbal instructions for each condition were simplified to provide clear explicit instructions for the participants to practice. The verbal instructions provided were as follows:

Split-attention condition:
To understand how to calculate angles $a$ and $b$, please do the following:
(1) Look at the diagram
(2) Read each text statement
(3) Re-read each text statement one at a time and search the diagram to see how the statement relates to the diagram and explains how to calculate the angle.
(4) Review the diagram and text statements to understand how each angle is calculated.
Integrated condition:
To understand how to calculate angles a and b, please do the following:
(1) Look at the diagram.
(2) Read each text statement next to the angle.
(3) Re-read each text statement, look at the diagram to understand how each angle is calculated.

Self-managed condition:
To understand how to calculate angles a and b, please do the following:
(1) Look at the diagram
(2) Read each text statement
(3) Re-read each text statement one at a time and move the text to a part of the diagram so it helps you understand how the angle is calculated.
(4) Review the diagram and text statements to understand how each angle is calculated.

Note how the moving of the text helps you to reduce your need to search between diagram and text.

After the two examples the researcher re-iterated to the participants that they were to use this learning technique in the four worked examples (Phase 3) they were about to learn.

Following the training for the instructional condition, an explanation of how to complete a subjective mental effort rating scale was provided. An established nine point Likert mental effort rating scale (Paas, 1992) was introduced. Participants were explained the concept of mental effort and provided with three examples to show the different extremities of the scale. Participants were instructed that they would be required to indicate the mental effort used on two occasions during the experiment, firstly after they had learnt the information and secondly after the completion of answering questions.

Phase 3: Learning to solve for angles on parallel lines (6 minutes)

After completing Phase 1 and Phase 2, participants worked through four worked examples, each worked example provided a four-step solution on how to solve alternate angle problems (worked examples 1 and 2) and co-interior angle problems (worked examples 3 and 4). Each worked example was presented on one Notebook slide (Figures 1-3 illustrate the first worked example). Participants were given 90 seconds on each slide and were then asked to move to the next slide. The researcher reiterated to the group of three students at each slide that they must use the learning approach they learnt in the training phase. The first and second worked examples showed how to solve for two angles on parallel lines using the rules; angles on a straight line add up to 180° and alternate angles are equal. The third and fourth worked examples required students to understand how to solve for two angles on a parallel line using the rules; angles on a straight line add up to 180° and co-interior angles add up to 180°. At the completion of the four slides participants were asked to circle the mental effort required to understand the information.

Following the learning phase, the researcher saved the Notebook files for all participants in the self-managed condition. Each file was saved and analyzed post-hoc to check that participants had implemented the guidance provided, i.e., compliance of the self-management condition. Participants who did not move text to some part of the diagram in an attempt to self-manage split-attention were not included in the final data analysis. Review of all Notebook artifacts showed one student failed to move text in the self-managed condition.

Phase 4: Transfer test (11 minutes)

Participants completed a paper-based test that included five parallel line angle problems with each problem presented on one page. Prior to starting the test, the participants were provided a worked example of how to answer test problems, where the researcher explained the need to provide the angle and a short justification on the reasons why the angle must be stated in degrees. Participants were required to complete questions in order, were not allowed to go back to a previous question, and had eleven minutes to complete all questions.

The test was designed to assess students’ ability to solve problems applying the three main rules learnt:
(1) Angles on a straight line add up to 180°.
(2) Identifying the position of alternate angles on parallel lines and applying the rule that alternate angles are equal.
(3) Identifying the position of co-interior angles on parallel lines and applying the rule that co-interior angles add up to 180°.

There were five test items, all of which required the students to transfer the acquired knowledge to a certain extent. The first two test items were similar to problems studied during the learning phase, but had different values. Participants needed to apply two rules to solve for two angles. The first item focused on solving for an alternate angle. The second item focused on solving for a co-interior angle. The last three test items were different to problems studied during the learning phase, and therefore required further transfer of the knowledge acquired during the learning phase. The third test item focused on solving for a co-interior angle but was presented differently to that studied during the learning phase in that the transversal line was presented in the opposite direction. The fourth test item required participants to solve for both alternate and co-interior angles and the last test item required solving for both alternate and co-interior angles on a problem that was both superficially and structurally different from what participants were exposed to in the learning phase (i.e., parallel lines were rotated 45 degrees).

Each problem was scored using a marking system developed by a mathematics teacher (one of the researchers – who also gave the introductory lesson). One mark was awarded for the correct angle e.g., \( r = 50° \) and one mark for the text written rationale, e.g., “angle r alternate to angle m, alternate angles are equal.” The first two test items comprised solving 2 angles; each were marked out of 4. The third test item comprised solving 2 angles and was marked out of 4. The fourth test item comprised solving 3 angles, and was marked out of six. The last test item included 4 angles to solve, and was marked out of eight. The maximum test score was 26.

Results

For all analyses a significance level of .05 was used. Cohen’s \( d \) was calculated as a measure of effect size, with values of .10, .30, and .50 characterizing small, medium, and large effect sizes, respectively (Cohen, 1988). One-way analyses of variance (ANOVAs) were conducted on transfer test performance scores and ratings of mental effort invested in the learning phase, and in the test phase. Means and standard deviations for transfer test performance and mental effort ratings in the learning and test phase are presented in Table 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Performance (Maximum score 26)</th>
<th>Mental effort ratings (Range: 1 – 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transfer test</td>
<td>Learning phase</td>
</tr>
<tr>
<td>Split-attention ((N = 15))</td>
<td>14.33 (6.40)</td>
<td>3.87 (1.60)</td>
</tr>
<tr>
<td>Integrated ((N = 17))</td>
<td>20.34 (5.73)</td>
<td>3.12 (1.50)</td>
</tr>
<tr>
<td>Self-managed ((N = 16))</td>
<td>19.44 (4.97)</td>
<td>3.50 (1.71)</td>
</tr>
</tbody>
</table>

Results from the one-way ANOVA for transfer-test performance indicated a significant main effect between the three conditions, \( F(2, 45) = 4.887, MSe = 32.630, p = .012 \). Post-hoc comparisons using the Tukey HSD test showed that the mean transfer test score of the self-managed condition was significantly higher than the mean score of the split-attention condition, \( p = .043, d = 0.89 \), indicating a large effect size (Cohen, 1988). Hypothesis 1 was thus confirmed. In addition, the mean transfer score of the integrated condition was significantly higher than the mean score of the split-attention condition, \( p = .015, d = 0.97 \), indicating a large effect size. The results replicate a split-attention effect thus confirming Hypothesis 2. There was no significant difference between the integrated and self-managed conditions, \( p = .915 \) and a low effect size was obtained, \( d = 0.15 \). This suggests that the self-managed condition performed equivalent to the integrated condition thus confirming Hypothesis 3.

Results from the one-way ANOVA for mental effort invested in the learning phase indicated no significant difference between the three conditions, \( F(2, 45) = 0.873, MSe = 2.567, p = .425 \). Thus Hypothesis 4 was not confirmed. However, a medium effect size was obtained between the integrated and split-attention conditions, \( d = 0.48 \); whilst a low effect size was obtained between the self-managed and split-attention conditions, \( d = 0.22 \), and between the integrated and self-managed conditions, \( d = 0.24 \).
Results from the one-way ANOVA for mental effort invested in the test phase indicated no significant difference between the three conditions, $F(2, 45) = 0.740$, $MSe = 3.355$, $p = .483$. Thus Hypothesis 5 was not confirmed. A medium effect size was obtained between the integrated and split-attention conditions, $d = 0.43$; a low effect size was obtained between the self-managed and split-attention conditions, $d = 0.21$, and between the integrated and self-managed conditions, $d = 0.21$.

Discussion

This study investigated the potential of students’ self-management of split-attention in worked examples for geometry learning in digital learning environment. The self-management strategy was compared to a condition in which students could learn from instructor-managed integrated worked examples, and to a condition in which students had to learn for split-attention worked examples. The results on transfer test performance confirmed Hypothesis 1, indicating that students who are provided the opportunity to physically integrate information sources in split-attention worked examples show better learning than students who study split-attention worked examples without the opportunity. In line with Hypothesis 2, we replicated the split-attention effect, indicating that students learn more from studying integrated worked examples than from studying split-attention worked examples. Hypothesis 3 was confirmed as the test performance results showed that there was no significant difference in performance between the integrated and self-managed conditions. Hypotheses 4 and 5 could not be confirmed, because the analyses revealed no significant differences between the instructional conditions for cognitive load during learning and during the test.

In the previous studies on the effects of self-management of split-attention on learning by Agostinho et al. (2013) and Roodenrys et al. (2012) the results revealed that learners’ adopting a self-management strategy to reduce split-attention performed better on test items compared to learners learning from split-attention instructions but learners did not perform as well as those learners provided with optimally designed instructions (integrated condition). Interestingly, the current study showed that the self-managed condition performed equivalent to the integrated condition. A possible reason for this different result in this current study is the fact that, in contrast to the previous studies, learners in the self-management condition clearly understood how and why they were required to integrate text statements within the diagram to support their understanding. It can be inferred from the results that this led to the success of the self-managed condition. Furthermore, the participants in the self-managed condition had the same learning time as the participants in the two other conditions, yet were required to undertake an additional task to meaningfully integrate text with diagram by moving the text boxes to the appropriate parts of the diagram. The results reveal that this additional task did not adversely affect learning.

The findings from this study suggest that firstly teaching learners how to self-manage the split-attention effect may lead to significantly better test results compared to learning from traditional split source instructions and secondly, allowing learners to self-manage the split-attention effect may be as effective as presenting learners with instructor-managed integrated instructions. These findings have an important implication, as learners today can access a variety of instructional materials, both print-based and online, and many of which may not be not optimally designed. Most importantly, when confronted with split-attention materials in a new learning tasks, learners who have been taught to self-manage split-attention materials can be expected to have an advantage over learners who have been taught with instructor-managed materials and do not know how to self-manage split-attention. This hypothesis needs to be tested in future research.

The results of this study have not only theoretical significance but also practical implications for learners’ studying mathematical instructions. Based on extensive research in cognitive load theory, instructor managed integrated instructions have been shown to be very effective in reducing split-attention and supporting learning (Chandler & Sweller, 1991; Tindall-Ford, Chandler, & Sweller, 1997; Ward & Sweller, 1990). In this study, the self-management strategy had a similar positive effect on learning. The result suggests that teaching learners to manipulate instructional material to reduce search and meaningfully integrate related sources of information may support their learning. There may be a number of factors in self-management of split-attention that result in positive learning outcomes, being: (1) the eventual reduction in search of the learners’ reorganized instructions; (2) the instructions being organized by the learner in a format that makes sense to them and; (3) the motivational factor in having control of instructions (Becker & Dwyer, 1994; Orvis, Fisher, & Wasserman, 2009)
There are several limitations to this study that could be addressed in future research. One limitation was that the school involved in the study was an all girls’ school. Previous research in self-management by Agostinho et al. (2013) and Roodenrys et al. (2012) included both male and female learners. Although, it seems unlikely that the positive effect found for the self-management strategy is gender specific, future research should replicate the study with male participants.

Another limitation of this study was that only the short-term benefit of teaching self-management of split-attention was examined. Future studies should investigate the long-term benefits of teaching self-management strategies, for example by including a delayed transfer task to examine if self-management strategies would be transferable to another context.

There are many avenues for future research in the area of self-management of cognitive load. One includes investigating self-management of other cognitive load effects, for example learners could be taught how to remove redundant information to support their learning. Another path could be examining the effects of movement as an alternative explanation for the positive effect of the self-managed condition. According to the theoretical framework of embodied cognition (Glenberg, 2008), one would expect the movement of digital objects to reduce cognitive load and result in richer cognitive schemata (see e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). Future studies could test this by comparing conditions in which students can self-manage with or without movements.

Overall, computer-based learning environments are increasingly offering media resources as discrete “objects” containing images, text, animations, and videos, along with tools enabling interactions with these objects through annotation features such as highlighting, typing notes, and other means of building associations between pieces or “objects” of information. The intention of this current research is to provide an evidence base to inform generalizable applications of declaring explicitly to learners the underlying strategies available to reduce extraneous activities such as search, and increasing germane activities associated with schema acquisition.

Conclusion

The research reported in this paper showed that learners who self-managed split-attention materials performed significantly better than learners who studied split-attention materials and performed as well as learners who studied instructor managed integrated instructions. Teaching learners how to self-manage split-attention is a new direction in CLT research. Today’s students are increasingly accessing a variety of online information, thus research that examines how learners can manipulate online information to better help them learn is an important endeavor.

Acknowledgements

This research was funded by the University of Wollongong’s Research Council Small Grant Scheme, 2013.

References


Learning from Concept Mapping and Hypertext: An Eye Tracking Study

Franck Amadieu1*, Ladislao Salmerón2, Julien Cegarra3, Pierre-Vincent Paubel1, Julie Lemarié1 and Aline Chevalier1

1CLLE research institute, University of Toulouse, France // 2University of Valencia, Valencia, Spain // 3ISAE, University of Toulouse, France // amadieu@univ-tlse2.fr // ladislao.salmeron@valencia.edu // julien.cegarra@univ-jfc.fr // paubel@univ-tlse2.fr // lemarie@univ-tlse2.fr // aline.chevalier@univ-tlse2.fr

*Corresponding author

ABSTRACT

This study examined the effects of prior domain knowledge and learning sequences on learning with concept mapping and hypertext. Participants either made a concept map in a first step and then read the hypertext’s contents combined with concept mapping (high activating condition), or they read the hypertext’s contents first and then made a concept map and re-read the hypertext’s contents (low activating condition). It was hypothesized that the low activating condition would support better learning of relations between concepts for low prior knowledge participants who would need information from hypertext first to efficiently build a map next. For high prior knowledge participants, it was expected that the high activating condition would increase prior knowledge activation that would improve learning by promoting germane cognitive load, or at least would help participants to cope with the cognitive demands of the learning task by reducing extraneous cognitive load. The results confirmed that the low activating condition fostered better learning of relations between the concepts than the high activating condition, regardless of the level of prior knowledge. However, concept mapping behaviors and eye movement data showed that prior knowledge reduced disorientation, improved navigation coherence, and supported better elaboration of semantic relations between the concepts before reading the texts.

Keywords

Concept mapping, Eye movements, Hypertext, Prior knowledge, Learning sequence

Introduction

Hypertexts are non-linear documents providing free access to different text sections by clicking on hyperlinks. Learning from hypertexts relies on comprehension processes that consist in establishing coherence between the consulted sections and constructing a mental representation of the overall semantic organization (Foltz, 1996). Selecting text sections by clicking on hyperlinks, maintaining coherent reading paths and extracting a structure of hypertext materials can be very demanding and imposes additional processes that could hamper the relevant processes of comprehension (Antonenko & Niederhauser, 2010; DeStefano & LeFevre, 2007; Niederhauser, Reynolds, Salmen, & Skolmoski, 2000). Pre-structured concept maps are relevant guiding tools for learners because they are explicit spatial overviews of the contents’ semantic organization and may thus limit learners’ disorientation (for a review see Amadieu & Salmerón, 2014).

While pre-structured concept maps limit some ineffective cognitive processes, they do not promote deep and active elaboration processes based on prior background knowledge. For instance, learners are less active and tend to navigate passively, respecting the structure imposed by the pre-structured map (Amadieu, van Gog, Paas, Tricot, & Mariné, 2009). Concept mapping can be seen as an alternative to pre-structured maps. Concept mapping requires that learners structure themselves a spatial overview of the semantic organization. Learners have to identify the main concepts from learning materials, and to organize them in a coherent manner by creating a spatial arrangement of concepts and links between the concepts to indicate the semantic relations. Coherence formation in learning from concept mapping is similar to coherence formation in learning from multiple representations, which requires the local structure of a single representation to be processed (intra-representational or local coherence formation) and to integrate the different representations (inter-representational or global coherence formation) to construct a deep understanding of the contents (Seufert, 2003). Global coherence formation is often considered as being an active learning task supporting elaboration (Karpicke & Blunt, 2011). Even though concept mapping is theoretically efficient for learning, empirical results are not clear cut (Hilbert & Renkl, 2009; Redford, Thiede, Wiley, & Griffin, 2012; Stull & Mayer, 2007) and the efficiency of this learning task seems to depend on several factors such as addition of heuristic examples (Hilbert & Renkl, 2009), or learners’ perceptions toward concept mapping (Tseng, Chang, Lou, Tan & Chiu, 2012). Because concept mapping frequently implies both building of a map and reading of textual material and because learners have to manage these two tasks, the temporal organization of the learning sequence including mapping and reading merits being studied. Therefore, designing an instruction that imposes a
specific sequence, i.e., processing textual material first vs. constructing a concept map first, could lead to different learning performance. This point was investigated in the current study in a hypertext-learning environment. A second factor considered in the present study was prior domain knowledge that can support both concept mapping and hypertext reading. The study investigated the on-line processes underpinning concept mapping with a hypertext.

Cognitive requirements of concept mapping: extraneous or germane cognitive processing?

Concept mapping may not be as efficient for learning as expected (e.g., Redford, Thiede, Wiley, & Griffin, 2012) and may even turn the learner away from learning (Stull & Mayer, 2007). Indeed, Stull and Mayer (2007) showed that concept mapping could impair content integration in contrast to studying already completed concept maps. The activity of constructing concept maps may induce extraneous processing, leaving less resources in working memory for essential processing. Different levels of cognitive requirements may be considered in concept mapping. Chang, Sung and Chen (2001) showed that filling in blank nodes and links in an incomplete concept map provided better learning than constructing an entire concept map. Gurlitt and Renkl (2008) also observed that, for high school students, creating and labeling links was more challenging and detrimental for learning than only labeling links. In another study, Gurlitt and Renkl (2009) found that when concept mapping was very demanding (i.e., creating links between concepts instead of just tagging already created links), learners performed worse in post-test measures of learning. These results argue for Cognitive Load Theory (Sweller, Ayres, & Kalyuga, 2011) considering that instructional design should avoid additional and ineffective cognitive processes (i.e., extraneous cognitive load) with regards to limited working memory resources.

Nevertheless, concept mapping might also contribute to active learning by fostering deep processing. In this line, the model of generative processes of comprehension (Wittrock, 1989), considers that relational processes (i.e., inferring relations between parts of a text and between text and prior knowledge) support deeper comprehension and learning. For instance, Bodemer, Plötzner, Feuerlein and Spada (2004) showed that learners who had to actively integrate different representations in a multimedia document by dragging and dropping the representations outperformed learners who received pre-integrated representations. As it contributes directly to the construction of meaning, relational processing required by concept mapping could be considered as belonging to the germane cognitive load. Learners who are able to produce elaborate concept maps would be engaged in germane cognitive processing as shown by Hilbert and Renkl (2008) who observed that learners who had good knowledge integration scores constructed more coherent concept maps. Moreover, Ponce and Mayer (2014) showed that filling in a graphic organizer close to a concept map led learners to conduct integrative strategies in their reading of a text.

In sum, instructional design would impose extraneous cognitive load if it makes relational processing difficult to run, or promote germane cognitive load if it makes it as a desirable difficulty (Bjork & Bjork, 2011). Leading learners to generate themselves connections between information parts may support germane cognitive load when learner’s prior knowledge level is appropriate.

Prior domain knowledge in concept mapping and hypertext reading

Prior domain knowledge plays a major role in the construction of meaning in comprehension (Kintsch, 1998) and learning (Sweller et al., 2011). Prior knowledge seems to contribute to process challenge imposed by concept mapping. Students with a high level of background knowledge can construct concept maps faster (Amadieu, Tricot, & Mariné, 2009) or more interconnected maps (Dogusoy-Taylan & Cagiltay, 2014). Actually, concept mapping can contribute to the activation of prior knowledge (Gurlitt & Renkl, 2008, 2009). Gurlitt and Renkl (2008) studied how different concept mapping tasks could activate prior knowledge before learning from a hypertext. Their results initially showed that concept mapping could be a relevant knowledge activation task that orients the processes in the following hypertext reading task (i.e., more focused and less explorative approach). As argued by the authors, concept mapping can contribute to activate prior knowledge and to help learners to identify what they already know and what they do not yet know. Next, they showed that inducing a low-coherent prior knowledge activation, consisting in the elaboration of high demanding concept mapping (i.e., creating and labeling links) was more advantageous for high prior knowledge learners. By contrast, a high-coherent prior knowledge activation that consisted in the elaboration of low demanding concept mapping (i.e., only labeling provided links) was more advantageous for low prior knowledge learners. Their findings are consistent with the idea that the mere existence of
background knowledge does not warrant those facilitative effects and that students need to be engaged in the task to activate their prior knowledge to contribute to germane cognitive load (for similar arguments in research on text and hypertext comprehension see Salmerón, Kintsch, & Cañas, 2006; Kintsch & Kintsch, 1995; McNamara, Kintsch, Songer, & Kintsch, 1996).

Text reading and concept mapping: which learning sequence?

Because concept mapping is often combined with a reading task, the type of complementarity between text reading and concept mapping (i.e., learning sequence) is an issue for designing effective learning instructions. In a recent study (Amadieu, Cegarra, Salmerón, Lemarié, Chevalier, & Paubel, 2013), it was observed that learners who spontaneously read several hypertext nodes before starting building the map tended to better comprehend the hypertext than learners who started earlier building the map. The study conducted by Hilbert and Renkl (2008) showed that a concept-mapping task after a reading task contributes to learning. However, building a concept map before reading texts may also have a positive effect on learning as shown by the study of Gurlitt and Renkl (2008). In line with this result, Bonestroo and De Jong (2012) showed that asking learners to create their own plan (i.e., organizing a sequence of concepts) before learning from a hypertext supported better knowledge acquisition of the concepts’ structure. No previous studies to our knowledge have compared what learning sequence (i.e., reading-before or mapping-before) is the most beneficial for learning. Therefore, further investigations into the effects of learning sequence and into the conditions of its efficiency are needed.

Overview of the present experiment and hypotheses

In the current study, the concept-mapping task consisted in spatially organizing the concepts on screen and in creating and labeling links between provided concepts. Two factors were investigated in the present study: the learning sequence including concept mapping and hypertext consultation (level of prior knowledge activation) and the learners’ prior domain knowledge. Two learning sequences were compared: a mapping-before condition considered as a condition of high activation of prior knowledge (i.e., in a first step, learners had to build a concept map with no access to the texts sections of the hypertext, and in a second step, learners could read the text sections and continue building the concept map) and a reading-before condition considered as a condition of lesser activation of prior knowledge (i.e., in a first step, learners had to read all text sections, and in a second step, they could reread the text sections and start building a concept map). Additionally, an originality of the current study concerns the use of eye-tracking to examine online processes during concept mapping (see also Dogusoy-Taylan & Cagiltay, 2014; Ponce & Mayer, 2014). Specifically, eye-tracking was used to assess two cognitive activities. First, students’ level of attention to the core concepts on the map was assessed by examining eye fixations (Lai et al., 2013). Dogusoy-Taylan and Cagiltay (2014) found, by using verbal protocols, that novices did not mention an initial strategy during the early steps of concept mapping, suggesting that they were less able to analyze the situation and to plan their map construction. As high prior knowledge individuals fixate on more areas that are relevant for the task (Cook, Wiebe, & Carter, 2008; Gegenfurtner, Lehtinen, & Säljö, 2011), the current study investigated the attention driven by prior knowledge by measuring the attention distribution on core concepts that reflects top-down processes. Second, the level of relational processing was assessed by measuring eye transitions between information sources (O’Keefe, Letourneau, Homer, Schwartz, & Plass, 2014; Ponce & Mayer, 2014). Yang, Chang, Chien, Chien and Tseng (2013) conducted analyses of saccade paths during multimedia learning and observed that high prior knowledge students conducted more inter-AOI (area of interest) scanning between multiple representations (i.e., saccade paths between the text and picture zones), indicating better integration of the different representations. The current study examined the learners’ inter-AOI scanning (i.e., transitions between concepts on the map).

It was hypothesized that low prior knowledge learners would benefit more from a low activating condition (i.e., reading-before sequence) than a high activating condition (i.e., mapping-before sequence). In contrast to a mapping-before condition, a reading-before condition provides relevant information from texts to learners early in the learning process, to build a rather coherent text representation. This representation should guide their following concept mapping activity by supporting global coherence formation. Therefore, for low prior knowledge participants a low activating condition should promote better learning of the relations between concepts (Hypothesis 1a), should limit extraneous cognitive load (Hypothesis 1b), should lead to more coherent navigation (Hypothesis 1c), should guide
attention on the most relevant map concepts (Hypothesis 1d) and should facilitate the processing of relevant relations between concepts (Hypothesis 1e).

As far as high prior knowledge learners are concerned, it was expected that building a concept map before reading a hypertext could promote a more intense activation of their prior knowledge (i.e., high activating condition). To structure a concept map before reading a hypertext, learners have to elaborate relations between provided concepts by mobilizing their background knowledge. In contrast, a reading-before task (i.e., low activating condition) should promote less activation of prior knowledge and lead learners to build a map more from text contents than from their prior knowledge. Therefore, it was hypothesized that high prior knowledge learners either benefit from a high activating condition or reach equivalent learning performance under both learning conditions because they are able to cope with concept-mapping early in learning (Hypothesis 2a). In a high activating condition, their cognitive load should be higher than in a low activating condition, but associated with better learning of relations between concepts (i.e., germane cognitive load) (Hypothesis 2b). Because more top-down processes should be activated in this condition, their navigation should be more coherent (Hypothesis 2c), they should pay more attention to the most relevant concepts on the map (Hypothesis 2d) and they should perform more relevant transitions between concepts (Hypothesis 2e).

Method

Participants

Seventy-five undergraduate psychology students volunteered to participate in the study (82.66 % of female participants; mean age = 19.31 years, $SD = 2.66$). Each participant received a purchase voucher of 15 euros for her/his participation at the end of the study. For the analyses, 10 participants were removed from the sample because of eye movement issues (unreliable calibration, loss of signal over the task). Hence, analyses were conducted on 65 participants.

Materials

Learning materials

A hypertext dealing with the greenhouse effect was designed. It consisted of twelve hypertext nodes corresponding to the main concepts (for a total of 635 words). A main page gathered together the twelve concepts, each being presented in a specific box. The concepts were linearly displayed on screen in an alphabetic order to avoid a coherent reading sequence. Participants could create the map on this main page by moving the concepts, drawing and labeling links between them (see Figure 1 for an example of concept map built by a participant). A label could be added to each created link by selecting one of 5 terms (i.e., belongs to, contributes to, emits, absorbs, reflects). Double-clicking on a link deleted it. Clicking on a concept opened the text dealing with this concept. The text appeared in full screen. A link below the text led back to the main page displaying the concept map where a new concept (or the same concept) could be opened. Therefore, participants could not simultaneously view the map and the texts on-screen. The participants were free to view the hypertext nodes several times and in any order.

Two prior knowledge activation conditions reflected by learning sequences were designed. For the high activating condition, during the first step of learning, participants were instructed to build a concept map from the boxes displayed on screen. The texts dealing with the concepts were not available during the first step. In the low activating condition, the participants were instructed to read all the texts, at least once each text, by clicking on the boxes organized in alphabetical order to access the corresponding text. They could neither move the concepts nor draw links between the concepts. When participants judged they had finished their map in the high activating condition, or had read enough the texts in the low activating condition, they were allowed to start the second learning step. During this second step, participants in the high activating condition could access to the texts by clicking on the links in the boxes (see Figure 1) and could still change their concept map, while participants in the low activating condition could build the map by moving the boxes and drawing links between them and could reread the texts. Participants were free to stop their learning for each step; nonetheless a time limitation of 30 minutes was imposed for each step (limitation assessed from preliminary tests).
Prior domain knowledge

According to the distinction between domain knowledge and topic knowledge (i.e., knowledge specific to the hypertext contents) proposed by Alexander, Kulikowich and Schulze (1994), prior domain knowledge was general knowledge about the subject matter, that is to say, knowledge of principles in physics and biology that are relevant to understand the greenhouse effect mechanism and climate change. An example of a question concerning the principle of absorption and emission of energy by objects is: “In physics, which principle is true?” Choices: (a) A physical object cannot absorb or emit energy, (b) A physical object can absorb energy without reemitting it, (c) A physical object can absorb and reemit energy (correct answer), (d) I do not know. Thirteen multiple-choice questions were used and validated by a physics teacher who taught the greenhouse effect. Each question had four possible choices including the answer “I do not know” to limit random answers. Each correct answer was awarded one point ($\alpha = .61$).

Measures of cognitive load

To measure the overall cognitive load, a scale was used (Paas, 1992): “The mental effort that you invested to study the mechanism of the greenhouse effect was.” “Mental effort is the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task; thus, it can be considered to reflect the actual cognitive load” (Paas, Tuovinen, Tabbers, & van Gerven, 2003, p. 64). In addition, to assess extraneous cognitive load linked to navigation, learners’ feelings of disorientation were measured by adaptation of a set of scales developed by Ahuja and Webster (2001). The scales were modified according to the material of the study: (a) “your difficulty knowing which text you had to view next was;”, (b) “your difficulty knowing where you were in the instructional document was;” and (c) “your difficulty finding information that you had previously read was;” ($\alpha = .76$). In both scales, a 7-point rating scale ($1 =$“Very low$, 7 =$“Very high”) was used.

Figure 1. Example of a concept map built by a participant
**Processing of concept mapping**

**Attention and relational inter-AOI scanning.** The areas of interest (AOI) were the boxes representing the concepts on the mapping page. Two indices from eye-tracking data were calculated: (1) Attention to core concepts was calculated as the number of fixations on the core concepts (i.e., the concepts most directly implicated in the greenhouse mechanism: Solar Radiation, Infrared Radiation, Ground, Greenhouse Gases including the concepts of Ozone, CO₂ and Water) divided by the number of fixations on all the map concepts (the number of fixations was divided by the number of words on the concept boxes to control word variability between the different concept boxes); (2) Relational inter-AOI scanning was calculated as the number of times eyes moved between concepts linked in the experts’ map (i.e., high-related pairs of concepts), divided by the number of times eyes moved between any concepts in the map. Thus, the index ranged from 0 to 1. This index indicated the degree to which students focused on the relationships between important map concepts, among all potential relationships between concepts on the map. Values closer to 1 indicated that the majority of visual transitions were relevant. In sum, we interpreted this index as a measure of relational deep-processing of the relations between important map concepts.

**Interconnectivity and quality of the concept maps**

The concept maps built by participants were taken as a measure of the quality of the mental model they elaborated. Two indices were calculated: (1) Number of links drawn between the concepts, indicating to what extent participants established connections between concepts; and (2) Relevance of the links. To calculate a relevance index, the map built by participants was compared to that produced by two high-school teachers in biology and physics. An expert’s concept map was produced, including all the links created by both experts. The relevant links created by participants corresponded to the links shared with the experts’ map. A relevance index of each participant’s map was calculated by dividing the number of drawn links similar to the experts’ map by all drawn links.

**Navigation coherence**

An index of navigation coherence was calculated as the number of navigation transitions between texts corresponding to a link of the experts’ map divided by the number of total navigation transitions (Amadieu, Tricot, & Mariné, 2010).

**Knowledge gain**

A pre-test and a post-test were used to calculate knowledge gain scores. To limit recognition of the questions by participants, the order of questions between the pre-test and the post-test was changed and five days were allowed to elapse between tests. Two types of knowledge were considered: (a) microstructural level (i.e., knowledge of explicit details specific to a concept mentioned within text sections) and (b) macrostructural level (i.e., knowledge of relations between concepts). The two types of knowledge were measured by multiple-choice questions (five choices per question including the answer “I do not know” to limit random answers) that are considered as relevant question formats to assess the mental model (Ozuru, Best, Bell, Witherspoon, & McNamara, 2007). Eight items measured the microstructural level (α = .46 for the pre-test & α = .60 for the post-test) and sixteen items measured the macrostructural level (α = .68 for the pre-test & α = .71 for the post-test).

**Apparatus**

Gaze data were recorded using an SMI RED 250 binocular eye tracker (SMI, Teltow, Germany). The sampling rate was set to 60Hz. This eye tracker has a spatial accuracy greater than 0.5°, and a 0.03° tracking resolution. A chin rest was used to maintain distance and to avoid large head movements. Eye movements for each participant were calibrated using five fixation points. A DELL 22” monitor with a refresh rate of 75 Hz and a resolution of 1680x1050 pixels was used. All fixations on areas of interest (i.e., concepts on the map) were detected using the SMI BeGaze default dispersion-based algorithm (set to 100 pixels and minimum duration of 80 msecs).
Procedure

During an initial session (approx. 40 minutes), participants performed the prior domain knowledge test followed by the pre-test specific to the greenhouse effect topic. They were then instructed how to use the program to open the texts from the concepts displayed on screen and how to construct a concept map by using the mouse. Thus, they practiced with the concept-mapping program to avoid unfamiliarity problems with the task and the functions of the program.

During the second session (approx. 50 minutes) conducted several days later, students performed the experimental task while their eye movements were recorded. Before starting each step, eye calibration was set up. They were randomly assigned to one of the learning conditions: high activating condition or low activating condition.

They were told that there were two learning steps and that they had to learn the greenhouse effect mechanisms to answer subsequent questions. In the high activating condition (i.e., mapping-before task), they were instructed to build a concept map from the provided concepts without texts to represent the greenhouse effect mechanism and were told that they would gain access to the texts and that they would be allowed to modify their map in the second step. In the low activating condition (i.e., reading-before task), participants were told they had to read the texts in the first step and, for the second step, that they would have to build a concept map and could access to the texts again. After that, they rated their mental effort (i.e., overall cognitive load) and perceived disorientation and performed the post-test questions.

Results

In order to test the effects of independent variables, prior knowledge activation (high activating vs. low activating) and prior knowledge, multiple regression analyses with interaction terms were conducted on each dependent variable. The independent variables were entered simultaneously into the regression model before entering the interaction term. Prior knowledge was entered as z-standardized variables. The level of prior domain knowledge was similar across conditions (for the low activating condition, $M = 6.94$, Min = 2, Max = 12, $SD = 2.23$, for the high activating condition, $M = 7.12$, Min = 3, Max = 12, $SD = 2.48$), $t(63) = 0.31$, $p = .76$. The learning condition was entered as a contrast-coded dummy variable.

First, the analyses of students’ learning performance and cognitive load measures are reported. Then, the analyses of learning processes including navigation and concept mapping processing are described.

Learning performance and cognitive load

Descriptive data of the knowledge gains and cognitive load measures are presented in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>High activating condition (Mapping-before task)</th>
<th>Low activating condition (Reading-before task)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gain of knowledge</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microstructure scores (max = 8)</td>
<td>1.73</td>
<td>1.38</td>
</tr>
<tr>
<td>Macrostucture scores (max = 16)</td>
<td>4.48</td>
<td>3.22</td>
</tr>
<tr>
<td><strong>Cognitive load measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental effort ratings (from 1 to 7)</td>
<td>4.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Disorientation ratings (from 1 to 7)</td>
<td>3.94</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Knowledge gains

Multiple regression analyses with the knowledge gain scores indicated no effect for the microstructure scores, $R_{corr}^2 = -.01$, $F(3, 61) = 0.73$, $p = .540$, though there was an effect for the macrostructure scores, $R_{corr}^2 = .09$, $F(3, 61) = 3.02$. 
A main effect of condition was observed, \( t(61) = 2.87, p < .006, \Delta R^2 = .117 \). Participants in the low activating condition outperformed participants in the high activating condition. There was no effect of prior knowledge, \( t(61) = 1.01, p = .316 \), and contrary to predictions, there was no interaction effect, \( t(61) = 0.30, p = .763 \). Thus, the reading-before task entailed better learning of the relations between the concepts, regardless of the level of prior knowledge.

Cognitive load: Perceived disorientation and overall cognitive load

The multiple regression analyses conducted on the perceived disorientation ratings showed a significant model, \( R_{corr}^2 = .10, F(3, 61) = 3.38, p = .024 \). Only prior domain knowledge had an effect, \( t(61) = -3.13, p < .003, \Delta R^2 = .138 \). The more prior domain knowledge participants had, the less disorientated they felt in the hypertext. There were no other effects (all \( p > .10 \)). By contrast, the multiple regression analysis of mental effort ratings reflecting the overall cognitive load did not indicate any effect, \( R_{corr}^2 = -.03, F(3, 61) = 0.37, p = .78 \).

Analyses of learning processes: Navigation, quality of maps and eye movements

Descriptive data of the navigation and concept mapping measures are presented in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Navigation, concept mapping and eye movements’ data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Rate of navigation coherence</td>
</tr>
<tr>
<td>Number of drawn links in the concept maps built by participants</td>
</tr>
<tr>
<td>Rate of relevant links in the concept maps built by participants</td>
</tr>
<tr>
<td>Rate of eye fixations on the core concepts</td>
</tr>
<tr>
<td>Rate of relational inter-AOI scanning</td>
</tr>
</tbody>
</table>

Navigation coherence

A multiple regression model indicated effects on navigation coherence for the second learning step (it was not possible to run the analysis on step 1 because participants in the high activating condition didn’t have to navigate), \( R_{corr}^2 = .14, F(3, 61) = 4.48, p = .007 \). There was a main positive effect of prior knowledge, \( t(61) = 2.85, p = .006, \Delta R^2 = .109 \). This result confirmed that high prior domain knowledge facilitated the selection of the next text to read according to its semantic proximity to the previously read text. Navigation coherence tended to be higher in the high activating condition than in the low activating condition, \( t(61) = -1.85, p = .070, \Delta R^2 = .046 \). Finally, there was no interaction, \( t(61) = -0.90, p = .372 \).

Interconnectivity and quality of concept maps

The examination of the drawn links on the participants’ maps at the end of the first learning step (only participants from the high activating condition could be considered) did not show any effect of prior knowledge, \( R_{corr}^2 = .00, F(1, 31) = 0.91, p = .342 \). However, the model was significant for the second step (including participants from both
conditions), \( R_{corr}^2 = .21, F(3, 61) = 6.77, p = .001 \). High prior knowledge facilitated the construction of more interconnected maps, \( t(61) = 2.48, p = .016 \), and the maps built in the high activating condition were more interconnected than in the low activating condition, \( t(61) = -3.62, p = .001 \). No interaction was observed, \( t(61) = -.32, p = .749 \).

Next, regarding the relevance of the maps, the initial map drawn in step 1 was analyzed first (only the high activating condition was concerned). Regression analyses revealed that prior knowledge led to a higher proportion of relevant links, \( R_{corr}^2 = .15, t(1, 31) = 2.56, p = .016 \). Second, the analyses of the final maps constructed in step 2 (including both conditions) did not show any significant model, \( R_{corr}^2 = .01, F(3, 61) = 1.11, p = .353 \).

**Attention to core concepts**

The rate of fixations on core concepts was analyzed. In the first step the model was significant, \( R_{corr}^2 = 0.08, F(3, 61) = 2.89, p = .043 \). Attention to core concepts was higher in the low activating condition than in the high activating condition, \( t(61) = 2.38, p = .021, \Delta R^2 = .081 \). There was no effect of prior knowledge, \( t(61) = 0.49 , p = .623 \), nor interaction, \( t(61) = -1.62, p = .110 \). No effect for the second step was observed, \( R_{corr}^2 = 0.02, F(3, 61) = 1.36, p = .264 \).

**Relational inter-AOI scanning**

For the first learning step, the analysis of the ratio of relational inter-AOI scanning revealed a significant model, \( R_{corr}^2 = .15, F(3, 61) = 4.66, p = .005 \). The ratio was significantly higher in the high activating condition than in the low activating condition, \( t(61) = -3.01, p = .004, \Delta R^2 = .121 \). Prior knowledge did not have any effect, \( t(61) = .85, p = .398 \). However the interaction tended to be significant, \( t(61) = -1.89, p = .064, \Delta R^2 = .047 \). Relational inter-AOI scanning increased with prior knowledge in the high activating condition, \( R_{corr}^2 = .22, F(1, 31) = 9.95, p = .004 \), but not in the low activating condition, \( R_{corr}^2 = -.02, F(1, 30) = 0.31, p = .579 \). For the second learning step, the analysis did not show any effect, \( R_{corr}^2 = -.05, F(3, 61) = 0.01, p = .99 \).

**Discussion**

This study investigated the effects of students’ prior domain knowledge and the level of prior knowledge activation induced by the learning sequence (i.e., reading-before task for the low activating condition vs. mapping-before task for the high activating condition) on learning with concept mapping from a hypertext. The results showed that the gain of the macrostructural level was higher when participants started by reading the hypertext than with concept mapping (Hypothesis 1a). Contrary to hypothesis 2a, this effect occurred whatever the level of prior knowledge. Microstructural level gains did not differ between the two learning conditions. This concurs with previous findings showing that concept mapping impacts global coherence formation and deep processing of material rather than retention (Stull & Mayer, 2007). To shed light on these unexpected results, we will next discuss how students’ prior knowledge influenced their concept mapping, as evidenced by building and eye-movement data.

**Students’ prior knowledge and concept map processing**

Eye data analyses indicated that the low activating condition led both high and low prior knowledge learners to distribute more attention to the core concepts during the first learning step (Hypothesis 1d) but not to pay more attention to the most relevant relations between concepts during mapping (Hypothesis 1e). Moreover, contrary to expectations, for the low prior knowledge learners, the low activating condition did not reduce the overall cognitive load (hypothesis 1b) and did not improve navigation (Hypothesis 1c). Concerning the high activating condition, the results indicated that building a map before reading a hypertext yielded more interconnected maps, regardless of the level of prior knowledge, but did not lead to higher map quality.

As far as high prior knowledge learners are concerned, the results showed that the low activating condition entailed better learning than the high activating condition as mentioned above (Hypothesis 2a). In addition, contrary to
expectations, for the high prior knowledge learners, the high activating condition did not support better navigation (Hypothesis 2c), or more attention paid to the core concepts (Hypothesis 2d), or more relational processes (Hypothesis 2e) during learning phase 2. Nevertheless, the study yielded interesting findings on the positive effects of prior knowledge on navigation and concept map processing. First, prior knowledge reduced extraneous cognitive load linked to navigation (i.e., disorientation ratings) and supported more coherent navigation corroborating previous findings (Amadieu et al., 2010). Second, prior knowledge supported the building of more interconnected maps as observed by Dogusoy-Taylan and Cagiltay (2014). These two results confirmed that prior knowledge contributed to more relational processing between the concepts. However, the clearest effects of prior knowledge on concept mapping were observed during the first learning step. In the high activating condition, prior knowledge led learners to build more coherent maps (i.e., better rate of relevant links between concepts). The examination of eye transitions between the concepts on the map confirmed that more attention was allocated to the most relevant relations between concepts. The results pertaining to the role of knowledge activation played by a mapping task before reading are consistent with previous findings (Bonestroo & De Jong, 2012; Gurlitt & Renkl, 2008). However, after text introduction, the effects of prior knowledge on the attention paid to the most relevant relations between concepts on the map vanished, showing no effect of knowledge activation on hypertext processing.

**Knowledge activation and learning from concept mapping**

Although the high activating condition supported prior knowledge activation during the first step, it did not favor learning for learners with high prior knowledge. Two main reasons can explain the lack of interaction. First of all, prior knowledge activation might also have occurred in the low activating condition. In this condition, learners read the texts in a non-coherent order because they were organized alphabetically. This could have concurred with prior knowledge activation by requesting strong inferential activity between texts to understand relations between them (i.e., the concepts) as shown in hypertexts (Amadieu et al., 2010) or text comprehension (McNamara et al., 1996). This explanation is supported by the results obtained on navigation, indicating that prior knowledge supported inferential activity in both conditions, as can be interpreted by their high levels of coherent navigation. The second reason that could explain the positive effect of the low activating condition on learning for both levels of prior knowledge concerns the function of the concept mapping task after reading the texts. During the first step, learners may have developed a rich text representation from the texts (i.e., text base, Kintsch, 1998) and may have sketched some interrelationships between concepts. The posterior concept-mapping task may have helped them to reflect on these relationships, which may in turn have boosted students’ inferential knowledge. Besides, reading the texts before could have help learners to plan their concept mapping. Indeed, Hilbert and Renkl (2008) found that a good planning predicted better performance in concept mapping.

In sum, the low activating condition could also have promoted inferential processing supporting better comprehension of relations between concepts (i.e., global coherence formation), but contrary to the high activating condition, the inferences drawn by learners would be more accurate or less ambiguous because they were conducted on the basis of both prior knowledge and text material. The texts fulfilled the function of aids by providing information about relations between the concepts that could have facilitated global coherence formation (Seufert & Brünen, 2006). Participants in the high activating condition may have kept in their mental representation incorrect relationships built during phase 1, even after reading the texts in phase 2, as indicated by a higher number of links on their maps. A potential means of testing this idea in future research would be to provide corrective feedback on the constructed maps before proceeding to the reading of the texts.

**Limitations and future research**

A potential limitation of the study is the knowledge background of participants. The lack of interaction of the learning conditions with prior knowledge may be due to the level of the participants who were psychology undergraduates with some notions of biology. Further studies should be conducted on students with more prior knowledge of the study materials. In line with this point, the use of a concept mapping task to activate students’ knowledge should be used with caution, or at least be restricted to the creation and labeling of links, as found in previous studies (Gurlitt & Renkl, 2008). Therefore, different levels of cognitive demands linked to concept mapping should be examined by comparing free concept mapping and guided concept mapping (e.g., labeling provided links). Another interpretation explaining that high prior knowledge learners benefited from the low activating condition is
that students may be more used, and probably may feel more confident with an activity that involves first reading a text and after doing something with it, such as writing a summary, answering questions, or building a concept map. A long-term study including several practice sessions with hypertext concept mapping may shed light on this point.

Although the current study provided a picture of the online processes engaged in concept mapping, questions remain over how each learning step contributed to learning. Thus, it would be informative to examine how comprehension and cognitive load evolves throughout a learning task involving text reading, concept mapping and both. Moreover, although the study showed that the low activating condition improved comprehension of the relations between concepts, the cognitive load measure results failed to prove that building a concept map before reading texts entailed more cognitive load. The lack of difference in disorientation ratings might be explained by the fact that the navigation task across the texts was quite similar between both learning conditions. Concerning the single retrospective cognitive load rating, as well as for the measures of comprehension, introducing intermediate measures of cognitive load after each learning step should provide information about the dynamic of cognitive demands over learning and should lead to different results than a single retrospective rating (Van Gog, Kirschner, Kester, & Paas, 2012). Investigations of the three types of cognitive load assessed by measures designed by Leppink, Paas, van der Vleuten, van Gog and van Merriënboer (2013) could help disentangle the different types of cognitive load. Moreover, to disentangle the types of cognitive processing consuming working memory resources linked to text reading and concept mapping, a dual task paradigm, using concurrent verbal and spatial tasks, could be used in future studies, as well as dual-task methods measuring executive control processes, such as the rhythm method recently tested by Park and Brünken (2015). The rhythm method provides an indicator for executive control by including inhibition processes and thereby may be considered as a good method of measure navigation through a hypertext. Besides, this continuous measure allows the assessment of fluctuations in cognitive load and cognitive load peaks.

In this study, the use of different online and offline methods to investigate relational processing engaged in a concept-mapping task in hypertext provided fruitful results. Relational processes to plan and build a concept map were captured, by eye movements (relational inter-AOI scanning) (Ponce & Mayer, 2014), the maps constructed by learners (number and relevance of drawn links in the concept map), along with the navigation paths (coherence of the transitions between node texts). While we obtained a complete picture of a complex learning task, future studies may include additional methods (e.g., verbal protocols, electroencephalography) to contribute to the investigation of learning.

**Acknowledgements**

The authors acknowledge the support of the French Research Agency (“Agence Nationale de la Recherche” - ANR) under reference ANR-11-JSH2-001 1.

The authors acknowledge the reviewers for their insightful and helpful comments on the paper.

**References**


Interactions Between Levels of Instructional Detail and Expertise When Learning with Computer Simulations

Yuling Hsu¹, Yuan Gao², Tzu-Chien Liu³* and John Sweller⁴

¹Institute of Education, Tzu Chi University, Taiwan // ²Smart Learning Institute, Beijing Normal University, China // ³Department of Educational Psychology and Counseling, National Taiwan Normal University, Taiwan // ⁴School of Education, University of New South Wales, Australia // eileenhsu2010@gmail.com // yuangao0304@gmail.com // tzuchien@ntnu.edu.tw // j.sweller@unsw.edu.au

*Corresponding author

ABSTRACT

Based on cognitive load theory, the effect of different levels of instructional detail and expertise in a simulation-based environment on learning about concepts of correlation was investigated. Separate versions of the learning environment were designed for the four experimental conditions which differed only with regard to the levels of written instructional detail. One hundred and forty Grade 10 (lower-expertise) and Grade 11 (higher-expertise) students participated in this experiment. In accord with the expertise reversal effect, the results supported the hypothesis that higher levels of instructional detail benefited learning for lower-expertise learners, whereas lower levels of detail facilitated learning for higher-expertise learners. It was concluded that the level of instructional guidance needed to match learners’ levels of expertise.

Keywords

Cognitive load theory, Expertise reversal effect, Simulation-based learning environment, Levels of instructional detail, Expertise levels

Introduction

In recent years, computer-assisted learning in a simulation-based environment has become increasingly available in many areas of education (e.g., Kolloffel, Eysink, de Jong, & Wilhelm, 2009; Liu, Lin, & Kinshuk, 2010; Morris, 2001; Renken & Nunez, 2013; Rutten, van Joolingen, & van der Veen, 2012; van der Meij & de Jong, 2006). Simulation-based learning attempts to model a real-life situation with dynamically linked multiple representations on a computer so that complex concepts can be visualized or modelled (Lee, Plass, & Homer, 2006; van der Meij & de Jong, 2006). Learning with computer simulations can be considered to have similarities with discovery learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Mayer, 2004). It provides a platform for learners to construct their own mental models about the concepts or knowledge to be learned by interacting with the environment.

Although the advantages of using simulation-based discovery learning have been confirmed by some empirical studies (e.g., Jaakkola, Nurmi, & Lehtinen, 2010; Lindgren & Schwartz, 2009; Urban-Woldron, 2009), many have argued that learning with minimal guidance (Kirschner, Sweller, & Clark, 2006) in a simulation-based environment often has proved to be ineffective (Eckhardt, Urhahne, Conrad, & Harms, 2013; Kanar & Bell, 2013; Mayer, 2004; Swaak & de Jong, 2001). Mayer (2004) reviewed a number of studies conducted from 1950 to the late 1980s comparing guided and unguided learning and suggested that learning was more effective using guided rather than unguided forms of instruction. In another study with two meta-analyses based on a sample of 164 research studies of discovery learning (Alfieri et al., 2011), the findings suggested that unassisted discovery failed to benefit learning unless the instructional supports in the form of feedback, work examples, scaffolding, and elicited explanations were provided.

In light of these results, how to design appropriate guidance for learning in simulation-based environments is critical (Rutten et al., 2012). Van der Meij and de Jong (2011) found that using step-by-step guidance for self-explanations to relate and translate between representations resulted in better learning outcomes than using general guidance in a simulation-based learning environment. Another study examining the sequential effects of high and low instructional guidance on children’s acquisition of experimentation skills within a discovery learning environment (Matlen & Klahr, 2013), indicated that learning and transfer were promoted whether high guidance instruction consisting of a combination of direct instruction and inquiry questions was received before or after low guidance that included inquiry questions only. A study by Lazonder and Egberink (2014) found that using segmented inquiry questions to scaffold learning procedures facilitated children’s acquisition and use of the control-of-variables strategy as much as
directly providing instruction prior to investigating a multivariable inquiry task in a simulation-based learning environment.

These studies indicated how learning guidance can be designed to improve learning processes and outcomes in a simulation-based environment; however, few studies have been conducted based on individual differences of learners. In studies based on cognitive load theory, optimal instructional designs have been found to interact with levels of expertise (Kalyuga, 2007; 2009a, 2009b; Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 2000, 2001). Cognitive load theory has been associated with learning with technology in recent years (e.g., De Koning, Tabbers, Rikers, & Paas, 2009; 2011; Lee et al., 2006; Liu, Lin, & Paas, 2013; Liu, Lin, Tsai, & Paas, 2012; Rey & Fischer, 2013).

Cognitive load theory

Cognitive load theory describes different sources of cognitive load when information is transited between working memory and long-term memory (Sweller, 1994, 2003, 2004, 2010, 2011, 2012; Sweller, Ayres, & Kalyuga, 2011). Three types of cognitive load are distinguished in cognitive load theory. Intrinsic cognitive load is generated by the level of complexity of the essential information involved in a learning task. Extraneous cognitive load is caused by an inappropriate format of instruction. Germane cognitive load refers to effective learning processes in which individuals use their cognitive resources to deal with the intrinsic cognitive load induced by learning tasks. Cognitive load theory suggests that instruction should be organized so that limited working memory resources are devoted to dealing with intrinsic and not extraneous cognitive load.

From the perspective of cognitive load theory, when learning with computer simulations, learners need to search, select, and manipulate related information from the multiple representations displayed in the simulation-based environment, which imposes an extraneous cognitive load because it is not directly related to learning and understanding. When the information is appropriately searched, selected, and manipulated, the learners are able to process and encode the selected information in working memory with the aid of prior knowledge stored in long-term memory. The processed information, integrated with existing knowledge in the form of schemas, must be stored in long-term memory. These processes are intrinsic to any learning task and so impose an intrinsic cognitive load whose level depends on the nature of the task. Cognitive load theory suggests that extraneous cognitive load should be minimized by optimal instructional design (Sweller et al., 2011). For this reason, in a simulation-based learning environment, learning how to search and select appropriate representations as well as how to manipulate the representations correctly is critical. If these processes are not provided with an adequate degree of guidance, novice learners will need to discover the processes themselves generating an extraneous cognitive load. If we vary instructional guidance indicating which representations should be selected for manipulation and how to manipulate them in an effective way, learners should experience different levels of cognitive load resulting in different learning outcomes. For less knowledgeable learners, more detailed instructional guidance on representation selection and manipulation should decrease extraneous cognitive load and benefit learning. Novices dealing with complex information must process many elements simultaneously in working memory due to the high element interactivity associated with a high intrinsic cognitive load (Sweller, 2010).

However, what constitutes element interactivity associated with cognitive load also depends on learners’ prior knowledge (Kalyuga, 2007, 2009a, 2009b; Kalyuga et al., 2000, 2001; Sweller, 2010). Experts can retrieve very complex, sophisticated schemas from long-term memory into working memory to assist understanding. Those schemas can act as a single element. As a consequence, high element interactivity for novices may not result in a high cognitive load for more knowledgeable learners if sufficient prior knowledge has been attained to treat many interacting elements as a single element for information processing (Gao, Low, Jin, & Sweller, 2013). The cognitive load caused by learning in a complex, simulation-based environment with less detailed instructional guidance may be tolerable for experts. Indeed, additional guidance providing more expert learners with information that they already have may have negative rather than positive consequences due to the redundancy effect (Sweller et al., 2011). Thus, additional guidance required by novices and so having a positive effect may be redundant for more expert learners resulting in a negative effect. This contrast leads to the expertise reversal effect with additional guidance being beneficial for novices but having negative consequences for more expert learners.
Overview of the present study

The present study investigated the effects of levels of instructional details when learning about concepts of correlation in a simulation-based learning environment for learners with different levels of expertise. Specifically, the consequences of guidance varying in levels of detail on both higher- and lower-expertise learners were explored. The level of instructional detail for the current experiment was determined by two factors related to learning in a simulation-based environment: (1) what representations (parameters) could be selected for manipulation within the simulation settings (dynamically linked multiple representations, e.g., “r value” and “paired x and y number table” in this study), and (2) how these representations should be manipulated for effective learning, reflecting the learning procedure. There were four conditions. (1) A High Detail Representation-High Detail Learning Procedure condition in which the learners explored the simulation-based learning environment with given values for setting the parameter representations and with given procedures. (2) A Low Detail Representation-High Detail Learning Procedure condition in which no exact value was provided for setting parameter representations and so the learners had to select appropriate values from a given range. However, the learners were still provided with given procedures to explore the simulation-based environment. (3) A High Detail Representation-Low Detail Learning Procedure condition in which the learners were provided with exact values for parameter representations, but they were not provided with detailed procedures for learning. (4) A Low Detail Representation-Low Detail Learning Procedure condition in which neither exact parameter values nor learning procedures were provided in detail.

The High Detail Representation-High Detail Learning Procedure condition provided the highest level of instructional guidance whereas the Low Detail Representation-Low Detail Learning Procedure condition provided the lowest level of guidance in the current study, with both the Low Detail Representation-High Detail Learning Procedure condition and the High Detail Representation-Low Detail Learning Procedure condition providing intermediate levels of guidance. Within the theoretical framework of cognitive load theory, an interaction between levels of instructional detail and learner expertise was hypothesized. More specifically, it was expected that the highly detailed learning guidance would be of benefit for learners with little expertise in the domain. With higher expertise, the benefits of more detailed instructional guidance were predicted to disappear or even reverse.

Method

Participants

One hundred and fifty-two (male = 97; female = 55) students participated in the current experiment. All students attended a public high school in northern Taiwan. Twelve (male = 3; female = 9) did not finish the whole experiment and therefore 140 participants were included in the final analyses period. Sixty-nine (male = 46; female = 23) were in grade 10 and had no tuition on the concepts of correlation in school before the experiment except some basic tuition about simple probability and statistical charts such as bar charts, pie charts, etc. These students were classified as lower-expertise, less knowledgeable students in the present study. In contrast, seventy-one participants (male = 44; female = 27) in grade 11 had learned the concepts of correlation in the final term of grade 10, and were classified as higher-expertise, more knowledgeable students. Within both expertise levels the participants were assigned to the four experimental conditions balanced according to their average scores of the previous two midterm mathematics exams.

Experiment environment

The experiment was conducted in the third version of the Simulation-Assistant Learning Statistic environment developed by Liu and his colleagues (Liu et al., 2010) to assist students to learn the basic concepts associated with correlation and revised based on the design of the present experiment and the results of a pilot study conducted one week before this experiment (See Figure 1 for a screenshot of the Simulation-Assistant Learning Statistic III). The Simulation-Assistant Learning Statistic III included two major areas in its interface: a learning guidance area on the left side and a dynamic linked multiple representation area on the right side. Except that the learning guidance varied between conditions, the environment was identical for all learners. The dynamically linked multiple representation area contained a blank for setting up and presenting the correlation coefficient value ($r$ value), a table for displaying and manipulating the two dimensions of each of the sample numbers on the X and Y axes, and a
scatter plot for showing the distribution of data points corresponding to the sample numbers. These three representations presented the concepts associated with correlation in different formats. If the students manipulated any of these representations, the corresponding changes would occur in the other representations simultaneously. All manipulation activities by the students in the current experiment were conducted in this dynamically linked multiple representations area. The learning guidance area included textual instructions varying between conditions. It instructed the learners to set appropriate values for the specific representation and to explore the concepts using appropriate procedures. Specifically, the High Detail Representation-High Detail Learning Procedure and the High Detail Representation-Low Detail Learning Procedure conditions had highly detailed guidance on setting values for specific representations. These values were provided with exact \( r \) values (e.g., input 0.8, 1, -0.8, and -1 in this study) and paired \( x, y \) values (e.g., five pairs of numbers) while the other conditions with low detailed guidance for the representation settings were only instructed to enter any number between a range of -1 and 1 in the blank for \( r \) value and to make any changes for the values of the paired \( x, y \) variables. An example of the learning guidance for the four conditions is presented in Appendix A. Similarly, the High Detail Representation-High Detail Learning Procedure and the Low Detail Representation-High Detail Learning Procedure conditions with highly detailed guidance on the way to explore the environment were provided with detailed steps for observing each relation between dynamically linked multiple representations, whereas the other conditions with low detailed guidance on learning procedures were only provided a general direction to explore the relations between dynamically linked multiple representations.

**Figure 1.** A screenshot of the simulation-assistant learning statistic

### Materials and procedure

The experimental materials used in the current experiment were compiled according to the Taiwan National Syllabus Program Requirements designed for Grade 10, Semester 2 mathematics. Three units with two tasks related to three major concepts associated with correlation were selected as the learning materials in the current study, including positive and negative correlations (Unit 1), correlation degree (Unit 2), and perfect correlations (Unit 3). Each unit contained two tasks in which the relations between the dynamically linked multiple representations in the simulation-based learning environment were explored separately by manipulating the \( r \) value (Task 1) and the values of the \( x \) and \( y \) variables (Task 2).
The experimental procedures involved a pre-experimental phase, a learning phase, and a test phase, being executed on two days with an interval of one week. On the first day, only the pre-experimental phase was scheduled, lasting fifteen minutes. On the second day, the learning phase (50 minutes) and the test phase (10 minutes) were administrated using a computer assigned to each participant (see Figure 2, for a schematic representation of the procedure).

![Figure 2. The procedure over the experiment](image)

**Pre-experimental Phase (on the first day)**
- Prior knowledge test
  - 8-item multiple choice questions

**Learning Phase (on the second day)**
- Prior instruction and practice before learning
- Condition 1: High Detail Representation-High Detail Learning Procedure
- Condition 2: Low Detail Representation-High Detail Learning Procedure
- Condition 3: High Detail Representation-Low Detail Learning Procedure
- Condition 4: Low Detail Representation-Low Detail Learning Procedure
- 3 units with 2 tasks learning
- (Immediately followed)

**Test Phase (on the second day)**
- Comprehension test
  - 8-item multiple-choice format questions
  - Cognitive load evaluation

**Pre-experimental phase**

A prior knowledge test was carried out one week before conducting the experiment in order to verify the classification of higher- and lower-expertise learners and also to ensure an equivalent initial understanding of the concepts of correlation between experimental conditions within expertise levels. The students were given fifteen minutes to answer eight multiple-choice questions on a sheet of paper. One point was given for a correctly answered question, and full marks for the prior knowledge test were therefore 8 points.

**Learning phase**

In the learning phase, every student was provided with a computer with the software of Simulation-Assistant Learning Statistic III installed. One page of instruction about basic knowledge of the key concepts associated with correlation (e.g., \(r\) value, sample size, category of correlation, etc.) was presented to the students on the computer screen for learning at the beginning of the learning phase. No time limit was required for basic knowledge acquisition, and the students could move to the next page when they finished reading.
After basic knowledge acquisition, the students were introduced to the representations involved in Simulation-Assistant Learning Statistic III as well as their functions by a three-minute video. The video showed how to manipulate the different representations, such as an \( r \) value and the values of the \( x, y \) variables. When the video ended, the students were allowed to practice manipulating the representations on their own for a limit of 1 minute. This time limit was adequate for students to finish their practice. The students were also given opportunities to ask questions after practice.

The students were then presented with the simulation-based learning environment with the learning guidance varying corresponding to their own experimental condition. They were presented with identical learning tasks and an identical simulation-based learning environment except that the amount of detail in the learning guidance varied according to the experimental condition.

**Test phase**

The test phase immediately followed the learning phase and was conducted on the same computer as used in the learning phase. The ten minute comprehension test consisted of 8 multiple-choice items (see Appendix B, for an example question of the comprehension test). The internal consistency of the comprehension tests was 0.61. At the end of the comprehension test, a nine-point Likert rating scale originally designed by Paas, Van Merrienboer and Adam (1994) was presented on the computer screen. Students were asked to indicate their cognitive load in completing the task, with 1-9 being “very very low,” “very low,” “low,” “slightly low,” “neither high nor low,” “slightly high,” “high,” “very high,” and “very very high,” respectively.

**Results**

Statistical significance for all tests was set at the .05 level except when otherwise indicated. Partial \( \eta^2 \) was used as the effect size index. Accordingly, .01, .06, and .14 are considered as the partial \( \eta^2 \) values reflecting small, medium and large effect sizes (Cohen, 1988).

**Prior knowledge test**

A 4 (Experimental Condition: High Detail Representation-High Detail Learning Procedure condition, High Detail Representation-Low Detail Learning Procedure condition, Low Detail Representation-High Detail Learning Procedure condition, Low Detail Representation-Low Detail Learning Procedure condition) × 2 (Expertise Level: lower-expertise learners, higher-expertise learners) analysis of variance (ANOVA) was conducted on the prior knowledge test accuracy scores (see Table 1, for the means and standard deviations of the test accuracy scores of the prior knowledge test). The results demonstrated a significant main effect of expertise level, \( F(1, 132) = 89.41, MSE = 2.51, p < 0.001 \), partial \( \eta^2 = .40 \), indicating that the division of learners into higher- and lower-expertise levels was valid. Neither a main effect of the experimental conditions, \( F(3, 132) = 1.95, MSE = 2.51, p = .13, \) nor an interaction between the experimental conditions and expertise levels was found, \( F(3, 132) = 1.62, MSE = 2.51. p = .19 \), verifying the equivalent initial understanding of the concepts of correlation between the experimental conditions within expertise levels. Based on these results, it was decided not to use the prior knowledge test scores in any other analyses of the present study.

**Table 1. Means and standard deviations on the prior knowledge test**

<table>
<thead>
<tr>
<th>Learner expertise</th>
<th>Conditions</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Detail Representation-High Detail Learning Procedure</td>
<td>Low Detail Representation-High Detail Learning Procedure</td>
<td>High Detail Representation-Low Detail Learning Procedure</td>
<td>Low Detail Representation-Low Detail Learning Procedure</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Lower</td>
<td>18</td>
<td>3.61</td>
<td>1.54</td>
<td>18</td>
<td>2.28</td>
</tr>
<tr>
<td>Higher</td>
<td>19</td>
<td>5.16</td>
<td>1.64</td>
<td>17</td>
<td>5.00</td>
</tr>
</tbody>
</table>
Comprehension test

Table 2 shows the means and standard deviations for the four experimental conditions on the comprehension test. We conducted a two-way analysis of covariance (ANCOVA), with four experimental conditions (High Detail Representation-High Detail Learning Procedure condition, High Detail Representation-Low Detail Learning Procedure condition, Low Detail Representation-High Detail Learning Procedure condition, Low Detail Representation-Low Detail Learning Procedure condition) and two expertise levels (lower-expertise learners, higher-expertise learners) as between-subjects factors. The average scores of the two previous midterm examinations in mathematics used to allocate students evenly to the four experimental conditions were used as a covariate.

Before analyzing the data using the two-way ANCOVA, the homogeneity of regression coefficients was tested. The results indicated that neither the interaction between the covariate variable and experimental conditions, $F(3, 130) = .55, p = .65$, nor the interaction between the covariate variable and expertise levels, $F(1, 130) = 1.60, p = .21$, was significant. Thus, neither homogeneity assumptions were violated.

The ANCOVA revealed no main effect of experimental conditions, $F(3, 131) = .72, MSE = 2.06, p = .54$. A significant main effect of expertise level was found, $F(1, 131) = 14.87, MSE = 2.06, p < .001$, partial $\eta^2 = .10$, indicating a better test performance for the higher-expertise learners than for the lower-expertise learners. A significant interaction between expertise level and experimental condition was also obtained, $F(3, 132) = 5.89, MSE = 2.06, p = .001$, partial $\eta^2 = .12$ (see Fig. 3., for a representation of the significant interaction).

Simple effects tests were used following the significant interaction on the comprehension test performance. For lower-expertise learners, a significant effect of experimental conditions on the comprehension test question accuracy scores was obtained, $F(3, 131) = 3.52, MSE = 2.06, p = .02$, partial $\eta^2 = .08$. Post hoc LSD tests showed that the High Detail Representation-High Detail Learning Procedure condition and the High Detail Representation-Low Detail Learning Procedure condition significantly outperformed the Low Detail Representation-Low Detail Learning Procedure condition. There were no other significant differences between experimental conditions.

For higher-expertise learners, there was also a significant effect of experimental condition on the comprehension test accuracy scores, $F(3, 131) = 3.08, MSE = 2.06, p = .03$, partial $\eta^2 = .07$. Post hoc LSD tests indicated the High Detail Representation-High Detail Learning Procedure condition performed significantly worse than any of the other conditions with no other significant differences.

For the cognitive load self-ratings during the comprehension test. The homogeneity of variance assumption was again tested before conducting the ANCOVA. The results indicated neither a significant interaction between the covariate variable and experimental conditions, $F(3, 130) = 1.22, p = .31$), nor a significant interaction between the covariate variable and expertise level, $F(1, 130) = 0.84, p = 0.36$, revealing a plausible assumption of homogeneity of regression coefficients. ANCOVA only revealed a main effect of expertise level, $F(1, 131) = 33.54, MSE = 3.50, p < .001$, partial $\eta^2 = .20$. Inspection of the means indicated that the lower-expertise learners found the test harder than the higher-expertise learners. The results failed to indicate either a main effect of experimental conditions, $F(3, 131) = 1.05, MSE = 3.50, p = .37$, or an interaction between expertise level and experimental condition, $F(3, 131) = .38, MSE = 3.50, p = .77$. 

Table 2. Means and standard deviations for the four experimental conditions on the comprehension test

<table>
<thead>
<tr>
<th>Variables</th>
<th>High Detail Representation-High Detail Learning Procedure</th>
<th>Low Detail Representation-High Detail Learning Procedure</th>
<th>High Detail Representation-Low Detail Learning Procedure</th>
<th>Low Detail Representation-Low Detail Learning Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$N$</td>
</tr>
<tr>
<td>Lower-expertise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>18</td>
<td>5.50</td>
<td>1.43</td>
<td>18</td>
</tr>
<tr>
<td>Cognitive load</td>
<td>18</td>
<td>6.17</td>
<td>1.98</td>
<td>18</td>
</tr>
<tr>
<td>Higher-expertise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>19</td>
<td>5.05</td>
<td>1.72</td>
<td>17</td>
</tr>
<tr>
<td>Cognitive load</td>
<td>19</td>
<td>4.11</td>
<td>1.76</td>
<td>17</td>
</tr>
</tbody>
</table>
Figure 3. The performance of comprehension test. Note. 1, 2, 3, and 4 on the abscissa axis (Conditions) represent: 1. High Detail Representation-High Detail Learning Procedure condition; 2. High Detail Representation-Low Detail Learning Procedure condition; 3. Low Detail Representation-High Detail Learning Procedure condition; 4. Low Detail Representation-Low Detail Learning Procedure condition.

Discussion

The aim of this study was to examine the effects of levels of instructional detail in supporting learners in a simulation-based environment. Four varying levels of details in guidance embedded in the same simulation-based learning environment were compared. It was expected that highly detailed guidance would be helpful for lower-expertise learners while lower levels of details in guidance would be beneficial for higher-expertise learners. Overall, the hypothesis was supported by a significant, disordinal interaction of test performance between experimental condition and expertise level on the comprehension test.

A simulation-based learning environment can provide learners with a platform in which learners’ own mental model or schemas can be constructed by independently interacting with the environment. In such a relatively free learning environment, determining what should be dealt with (representation selection) in the environment and how to deal with the selected representation (learning procedure) is critical for effective learning (Kirschner, Sweller, & Clark, 2006). From a cognitive load theory perspective, free exploration of a highly complex environment may result in heavy demands on working memory (Sweller et al., 2011). If minimal details in guidance are provided, learners without sufficient prior knowledge need to devote a large amount of their working memory resources to search for problem solutions, leaving few working memory resources available for learning. The present results supported this hypothesis by demonstrating a positive learning effect using more detailed guidance when learners’ expertise levels were low. Although the High Detail Representation-High Detail Learning Procedure condition did not show a statistically significant superiority on comprehension test performance over the two conditions with intermediate levels of detailed guidance (High Detail Representation-Low Detail Learning Procedure condition & Low Detail Representation-High Detail Learning Procedure condition), the means revealed the expected trend using lower-expertise learners. A significant difference was found between the High Detail Representation-High Detail Learning Procedure condition and the Low Detail Representation-Low Detail Learning Procedure condition. The benefit of more detailed guidance on learning in a simulation-based learning environment was further confirmed by the significantly better comprehension test performance for the High Detail Representation-Low Detail Learning Procedure condition than for the Low Detail Representation-Low Detail Learning Procedure condition.

However, such learning effects due to highly detailed guidance were absent for higher-expertise learners. On the one hand, more knowledgeable learners had sufficient prior knowledge to process multiple elements as a single element when searching and selecting representations, resulting in the cognitive load associated with exploring a complex simulation-based environment being manageable. For example, higher-expertise learners previously had learned the
concepts of correlation and they were likely to know which manipulations would be most beneficial to further advance their understanding. Learning may therefore have been facilitated using a relatively free mode of exploration with few details during guidance. In contrast, highly detailed guidance that provided exact values for setting parameter representations and specific procedures for manipulating the representations already were likely to be part of the prior knowledge base of higher-expertise learners. The cognitive load associated with reading and following guidance was therefore redundant and imposed an extraneous cognitive load for higher-expertise learners, impairing learning.

These results are consistent with previous cognitive load theory studies on worked examples (Kalyuga et al., 2001; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Sweller, Van Merrienboer, & Paas, 1998; Tuovinen & Sweller, 1999) that found that worked-out examples instruction, which was effective for novices, lost its advantage and became redundant as the knowledge level of learners was raised as a consequence of intensive training. For example, Kalyuga et al. (2001) presented either a series of worked examples or a less guided, exploratory-based environment to mechanical trade apprentices. The results showed that when dealing with complex tasks, inexperienced trainees clearly benefited most from the worked examples procedure but the advantage disappeared after two training sessions as participants became more experienced in the domain. Despite the differences in procedure, the current results may be interpreted in the same way as that of the worked example studies: discovery with minimal guidance imposes a heavy cognitive load on less knowledgeable learners, limiting effective learning, especially in complex simulation-based learning environment. The disadvantages of discovery learning disappeared and reversed with higher expertise. These results provide a clear example of the expertise reversal effect (Kalyuga, 2007, 2009a, 2009b; Kalyuga et al., 2003), demonstrating a reversal of cognitive load effects with expertise. The present study, by examining the effects of instructional guidance when exploring a simulation-based environment, demonstrated that whether the instructional guidance imposed a germane or extraneous load on learners was determined by learners’ expertise.

One limitation of the present study lies in the cognitive load evaluations. On the comprehension test, except for the significant difference of the cognitive load self-ratings between higher- and lower-expertise learners, no other effects related to subjective measures of cognitive load were found. The failure to find subjective ratings effects was inconsistent with the test performance where differences between conditions were found. A possible reason for the absent main effects of subjective ratings of cognitive load was that the complex simulation technology imposed some cognitive load by itself and learners may have found it quite hard to differentiate how much cognitive load was imposed by the learning tasks rather than from dealing with the technology. This limitation in cognitive load measurement may be a challenge for all studies investigating technology-assisted learning. Future research on simulation-based learning environment may need to measure the cognitive load by more specific questionnaires that, for example, ask students to rate how much mental effort was spent on the learning tasks separately from how much mental effort was spent on using the technology. Alternatively, repeatedly measuring mental effort after each question of a test may provide a more accurate measurement of cognitive load (Van Gog, Kirschner, Kester, & Paas, 2012).

The relatively low internal consistency for the comprehension test (.61) constituted a second limitation of the present study. The small number of test questions is a possible reason for the reduced Cronbach’s alpha value (Nunnally & Bernstein, 1994). Although the comprehension test questions had been used in previous research with high internal consistency values (Liu, Kinshuk, Lin, & Wang, 2012; Liu et al., 2010), this study only selected eight questions from the previous tests. More test questions should lead to higher internal consistency values.

Another limitation of the present study was its failure to clearly distinguish between the Low Detail Representation-High Detail Learning Procedure condition and the High Detail Representation-Low Detail Learning Procedure condition, which had highly detailed guidance on either representation selection or learning procedures. As a consequence, it is unknown whether representations selection or learning procedure is more important in instructional guidance. Future research may explore this question by further analyzing the cognitive processes during learning, for example, by analyzing the learning performance recorded in the log file or measuring the cognitive load during the learning stage.

In conclusion, individual differences in learner expertise in a domain should be taken into account when selecting user-adapted levels of guidance for simulation-based discovery learning. For novice learners without sufficient prior knowledge, highly detailed guidance should be presented first. After learners become more knowledgeable in the
domain, the guidance can be less detailed in line with the guidance fading effect (Renkl & Atkinson, 2003; Renkl, Atkinson, & Große, 2004; Renkl, Atkinson, & Maier, 2000; Renkl, Atkinson, Maier, & Staley, 2002). These findings may provide a reason for the diverse results in previous research on the effects of guidance for simulation-based learning. For example, van der Meij and de Jong (2011) found a positive effect of learning guidance during computer simulation while Chang, Chen, Lin, and Sung (2008) obtained a negative effect. These apparent contradictions may be explained by the current findings.

Acknowledgments

The authors would like to thank the assistance of Wang-Shen Lu in developing the experiment environment. Moreover, we would like to thank the Ministry of Science and Technology, Taiwan, R.O.C. for financially supporting this research under Grant no. MOST 101-2511-S-003 -061 -MY3, and MOST 103-2511-S-003 -058 -MY3. Finally, we would like to thank the support of the “Aim for the Top University Project” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 104-2911-I-003-301.

References


Appendix A

An example of the learning guidance for the four experimental conditions with different levels of detail

**Unit 1: The concepts of positive and negative correlations**

The default sample size is 20 in this study \((n = 20)\).

Please complete tasks 1 & 2 using the following procedures.

At the same time, please observe the changes that occur in the \(r\) values, paired \(x\) and \(y\) values, and data points of the scatter plot as well as the relations between them.

### Learn about positive and negative correlations by setting different correlation coefficients (\(r\) values)

<table>
<thead>
<tr>
<th>Condition 1: High Detail Representation-High Detail Learning Procedure</th>
<th>Condition 2: Low Detail Representation-High Detail Learning Procedure</th>
<th>Condition 3: High Detail Representation-Low Detail Learning Procedure</th>
<th>Condition 4: Low Detail Representation-Low Detail Learning Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations Procedure</td>
<td>Given (r) value</td>
<td>No given (r) value</td>
<td>Given (r) value</td>
</tr>
<tr>
<td>Step 1</td>
<td>Set the (r) value to 0.7</td>
<td>Set the (r) value to any (r) value ranging from -1 to 1.</td>
<td>Try to observe the changes and the relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot when setting the (r) value to 0.7, -0.7 as well as pressing the reordering key on the (x), (y) table. Once again, reset the (r) values to 0.9 and -0.9, and observe the changes and relations.</td>
</tr>
<tr>
<td>Step 2</td>
<td>Press the reordering key on the (x), (y) table and observe the relations between the paired (x), (y) values and the (r) value.</td>
<td>Observe the relations between the distribution of data points on the scatter plot and (r) values.</td>
<td>Observe the relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot.</td>
</tr>
<tr>
<td>Step 3</td>
<td>Observe the relations between the distribution of data points on the scatter plot and (r) values.</td>
<td>Observe the relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot.</td>
<td>Observe the relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot.</td>
</tr>
<tr>
<td>Step 4</td>
<td>Input six pairs of (x), (y) values that are positively correlated as shown in the following table.</td>
<td>Input any six or more pairs of (x), (y) values that are positively correlated.</td>
<td>Input any six or more pairs of (x), (y) values that are positively correlated.</td>
</tr>
<tr>
<td>Step 5</td>
<td>According to the same procedure as above, reset the (r) values to 0.9 and -0.9, and observe the relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot.</td>
<td>According to the same procedure as above, reset the (r) values to another pair of inverse numbers ranging from -1 to 1, and observe the relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot.</td>
<td>According to the same procedure as above, reset the (r) values to another pair of inverse numbers ranging from -1 to 1, and observe the relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot.</td>
</tr>
</tbody>
</table>

### Task 2: Learn about positive and negative correlations by setting different paired \(x\) and \(y\) values

<table>
<thead>
<tr>
<th>Condition 1: High Detail Representation-High Detail Learning Procedure</th>
<th>Condition 2: Low Detail Representation-High Detail Learning Procedure</th>
<th>Condition 3: High Detail Representation-Low Detail Learning Procedure</th>
<th>Condition 4: Low Detail Representation-Low Detail Learning Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations Procedure</td>
<td>Given paired (x), (y) values</td>
<td>No given paired (x), (y) values</td>
<td>Given paired (x), (y) values</td>
</tr>
<tr>
<td>Step 1</td>
<td>Press “clear” to remove all the data.</td>
<td>After pressing “clear” to remove all the data, try to observe the changes and relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot by</td>
<td>After pressing “clear” to remove all the data, try to observe the changes and relations between the (r) values, paired (x) and (y) values, and data points of the scatter plot by</td>
</tr>
<tr>
<td>Step 2</td>
<td>Input six pairs of (x), (y) values that are positively correlated as shown in the following table.</td>
<td>Input any six or more pairs of (x), (y) values that are positively correlated.</td>
<td>Input any six or more pairs of (x), (y) values that are positively correlated.</td>
</tr>
</tbody>
</table>
inputting five pairs of x, y values with a positive correlation and then another five pairs of x, y values with a negative correlation as shown in the following tables.

<table>
<thead>
<tr>
<th>Positive</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>48</td>
<td></td>
</tr>
</tbody>
</table>

Step 3
Observe the ranking orders of x and y values (whether the x and y values increase or decrease).

Step 4
Observe the changes of r values and think about the relations between the paired x, y values and an r value.

Step 5
Observe the relations between paired x, y values and the distribution characteristics of data points on the scatter plot.

Step 6
Observe the relations between the r value and the distribution characteristics of data points on the scatter plot.

Step 7
Observe the relations between the r values, paired x and y values, and data points of the scatter plot.

Step 8
Press “update” to obtain another six pairs of x, y values with the same r value, and repeat procedures 3-7.

Step 7
Using the same procedure as above, input five new pairs of x, y values with a negative correlation as shown in the following table and observe the changes and relations between the r values, paired x and y values, and data points of the scatter plot.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>68</td>
</tr>
<tr>
<td>30</td>
<td>59</td>
</tr>
<tr>
<td>40</td>
<td>57</td>
</tr>
<tr>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>60</td>
<td>50</td>
</tr>
</tbody>
</table>

According to the same procedure as above, input six or more new pairs of x, y values with a negative correlation, and observe the changes and relations between the r values, paired x and y values, and data points of the scatter plot.

<table>
<thead>
<tr>
<th>Negative</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

inputting any five or more pairs of x, y values with a positive correlation and then repeat with a negative correlation.
Appendix B

An example question from the comprehension test

5. A researcher using a sample size of 9 collected data about how much students’ read literature and their literature test results in order to find the correlation between the two variables. The results indicated a positive correlation but not a perfect positive correlation between the amount of literature reading (X) and the literature test results (Y). Can you please indicate which of the following four sets of data about X and Y represent the data collected in this research?

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>23</td>
<td>32</td>
<td>68</td>
<td>67</td>
</tr>
<tr>
<td>Y</td>
<td>23</td>
<td>32</td>
<td>72</td>
<td>71</td>
</tr>
<tr>
<td>X</td>
<td>31</td>
<td>33</td>
<td>69</td>
<td>71</td>
</tr>
<tr>
<td>Y</td>
<td>42</td>
<td>49</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>X</td>
<td>41</td>
<td>43</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Y</td>
<td>53</td>
<td>51</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>X</td>
<td>48</td>
<td>48</td>
<td>51</td>
<td>84</td>
</tr>
<tr>
<td>Y</td>
<td>62</td>
<td>62</td>
<td>52</td>
<td>91</td>
</tr>
<tr>
<td>X</td>
<td>52</td>
<td>52</td>
<td>55</td>
<td>92</td>
</tr>
<tr>
<td>Y</td>
<td>61</td>
<td>61</td>
<td>52</td>
<td>99</td>
</tr>
<tr>
<td>X</td>
<td>67</td>
<td>67</td>
<td>76</td>
<td>77</td>
</tr>
<tr>
<td>Y</td>
<td>67</td>
<td>67</td>
<td>84</td>
<td>81</td>
</tr>
<tr>
<td>X</td>
<td>71</td>
<td>72</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>Y</td>
<td>72</td>
<td>72</td>
<td>96</td>
<td>91</td>
</tr>
<tr>
<td>X</td>
<td>72</td>
<td>72</td>
<td>94</td>
<td>93</td>
</tr>
<tr>
<td>Y</td>
<td>72</td>
<td>72</td>
<td>96</td>
<td>99</td>
</tr>
</tbody>
</table>

A. I, IV  
B. I, II, III, IV  
C. II, III  
D. I  
E. IV
Interactions Between Levels of Instructional Detail and Expertise When Learning with Computer Simulations

Yuling Hsu1, Yuan Gao2, Tzu-Chien Liu3* and John Sweller4

1Institute of Education, Tzu Chi University, Taiwan // 2Smart Learning Institute, Beijing Normal University, China // 3Department of Educational Psychology and Counseling, National Taiwan Normal University, Taiwan // 4School of Education, University of New South Wales, Australia // eileenhsu2010@gmail.com // yuangao0304@gmail.com // tzuchien@ntnu.edu.tw // j.sweller@unsw.edu.au

*Corresponding author

ABSTRACT

Based on cognitive load theory, the effect of different levels of instructional detail and expertise in a simulation-based environment on learning about concepts of correlation was investigated. Separate versions of the learning environment were designed for the four experimental conditions which differed only with regard to the levels of written instructional detail. One hundred and forty Grade 10 (lower-expertise) and Grade 11 (higher-expertise) students participated in this experiment. In accord with the expertise reversal effect, the results supported the hypothesis that higher levels of instructional detail benefited learning for lower-expertise learners, whereas lower levels of detail facilitated learning for higher-expertise learners. It was concluded that the level of instructional guidance needed to match learners’ levels of expertise.

Keywords

Cognitive load theory, Expertise reversal effect, Simulation-based learning environment, Levels of instructional detail, Expertise levels

Introduction

In recent years, computer-assisted learning in a simulation-based environment has become increasingly available in many areas of education (e.g., Kolloffel, Eysink, de Jong, & Wilhelm, 2009; Liu, Lin, & Kinshuk, 2010; Morris, 2001; Renken & Nunez, 2013; Rutten, van Joosten, & van der Veen, 2012; van der Meij & de Jong, 2006). Simulation-based learning attempts to model a real-life situation with dynamically linked multiple representations on a computer so that complex concepts can be visualized or modelled (Lee, Plass, & Homer, 2006; van der Meij & de Jong, 2006). Learning with computer simulations can be considered to have similarities with discovery learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Mayer, 2004). It provides a platform for learners to construct their own mental models about the concepts or knowledge to be learned by interacting with the environment.

Although the advantages of using simulation-based discovery learning have been confirmed by some empirical studies (e.g., Jaakkola, Nurmi, & Lehtinen, 2010; Lindgren & Schwartz, 2009; Urban-Woldron, 2009), many have argued that learning with minimal guidance (Kirschner, Sweller, & Clark, 2006) in a simulation-based environment often has proved to be ineffective (Eckhardt, Urahne, Conrad, & Harms, 2013; Kanar & Bell, 2013; Mayer, 2004; Swaak & de Jong, 2001). Mayer (2004) reviewed a number of studies conducted from 1950 to the late 1980s comparing guided and unguided learning and suggested that learning was more effective using guided rather than unguided forms of instruction. In another study with two meta-analyses based on a sample of 164 research studies of discovery learning (Alfieri et al., 2011), the findings suggested that unassisted discovery failed to benefit learning unless the instructional supports in the form of feedback, work examples, scaffolding, and elicited explanations were provided.

In light of these results, how to design appropriate guidance for learning in simulation-based environments is critical (Rutten et al., 2012). Van der Meij and de Jong (2011) found that using step-by-step guidance for self-explanations to relate and translate between representations resulted in better learning outcomes than using general guidance in a simulation-based learning environment. Another study examining the sequential effects of high and low instructional guidance on children’s acquisition of experimentation skills within a discovery learning environment (Matlen & Klahr, 2013), indicated that learning and transfer were promoted whether high guidance instruction consisting of a combination of direct instruction and inquiry questions was received before or after low guidance that included inquiry questions only. A study by Lazonder and Egberink (2014) found that using segmented inquiry questions to scaffold learning procedures facilitated children’s acquisition and use of the control-of-variables strategy as much as
directly providing instruction prior to investigating a multivariable inquiry task in a simulation-based learning environment.

These studies indicated how learning guidance can be designed to improve learning processes and outcomes in a simulation-based environment; however, few studies have been conducted based on individual differences of learners. In studies based on cognitive load theory, optimal instructional designs have been found to interact with levels of expertise (Kalyuga, 2007; 2009a, 2009b; Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 2000, 2001). Cognitive load theory has been associated with learning with technology in recent years (e.g., De Koning, Tabbers, Rikers, & Paas, 2009; 2011; Lee et al., 2006; Liu, Lin, & Paas, 2013; Liu, Lin, Tsai, & Paas, 2012; Rey & Fischer, 2013).

**Cognitive load theory**

Cognitive load theory describes different sources of cognitive load when information is transited between working memory and long-term memory (Sweller, 1994, 2003, 2004, 2010, 2011, 2012; Sweller, Ayres, & Kalyuga, 2011). Three types of cognitive load are distinguished in cognitive load theory. Intrinsic cognitive load is generated by the level of complexity of the essential information involved in a learning task. Extraneous cognitive load is caused by an inappropriate format of instruction. Germane cognitive load refers to effective learning processes in which individuals use their cognitive resources to deal with the intrinsic cognitive load induced by learning tasks. Cognitive load theory suggests that instruction should be organized so that limited working memory resources are devoted to dealing with intrinsic and not extraneous cognitive load.

From the perspective of cognitive load theory, when learning with computer simulations, learners need to search, select, and manipulate related information from the multiple representations displayed in the simulation-based environment, which imposes an extraneous cognitive load because it is not directly related to learning and understanding. When the information is appropriately searched, selected, and manipulated, the learners are able to process and encode the selected information in working memory with the aid of prior knowledge stored in long-term memory. The processed information, integrated with existing knowledge in the form of schemas, must be stored in long-term memory. These processes are intrinsic to any learning task and so impose an intrinsic cognitive load whose level depends on the nature of the task. Cognitive load theory suggests that extraneous cognitive load should be minimized by optimal instructional design (Sweller et al., 2011). For this reason, in a simulation-based learning environment, learning how to search and select appropriate representations as well as how to manipulate the representations correctly is critical. If these processes are not provided with an adequate degree of guidance, novice learners will need to discover the processes themselves generating an extraneous cognitive load. If we vary instructional guidance indicating which representations should be selected for manipulation and how to manipulate them in an effective way, learners should experience different levels of cognitive load resulting in different learning outcomes. For less knowledgeable learners, more detailed instructional guidance on representation selection and manipulation should decrease extraneous cognitive load and benefit learning. Novices dealing with complex information must process many elements simultaneously in working memory due to the high element interactivity associated with a high intrinsic cognitive load (Sweller, 2010).

However, what constitutes element interactivity associated with cognitive load also depends on learners’ prior knowledge (Kalyuga, 2007, 2009a, 2009b; Kalyuga et al., 2000, 2001; Sweller, 2010). Experts can retrieve very complex, sophisticated schemas from long-term memory into working memory to assist understanding. Those schemas can act as a single element. As a consequence, high element interactivity for novices may not result in a high cognitive load for more knowledgeable learners if sufficient prior knowledge has been attained to treat many interacting elements as a single element for information processing (Gao, Low, Jin, & Sweller, 2013). The cognitive load caused by learning in a complex, simulation-based environment with less detailed instructional guidance may be tolerable for experts. Indeed, additional guidance providing more expert learners with information that they already have may have negative rather than positive consequences due to the redundancy effect (Sweller et al., 2011). Thus, additional guidance required by novices and so having a positive effect may be redundant for more expert learners resulting in a negative effect. This contrast leads to the expertise reversal effect with additional guidance being beneficial for novices but having negative consequences for more expert learners.

114
Overview of the present study

The present study investigated the effects of levels of instructional details when learning about concepts of correlation in a simulation-based learning environment for learners with different levels of expertise. Specifically, the consequences of guidance varying in levels of detail on both higher- and lower-expertise learners were explored. The level of instructional detail for the current experiment was determined by two factors related to learning in a simulation-based environment: (1) what representations (parameters) could be selected for manipulation within the simulation settings (dynamically linked multiple representations, e.g., “r value” and “paired x and y number table” in this study), and (2) how these representations should be manipulated for effective learning, reflecting the learning procedure. There were four conditions. (1) A High Detail Representation-High Detail Learning Procedure condition in which the learners explored the simulation-based learning environment with given values for setting the parameter representations and with given procedures. (2) A Low Detail Representation-High Detail Learning Procedure condition in which no exact value was provided for setting parameter representations and so the learners had to select appropriate values from a given range. However, the learners were still provided with given procedures to explore the simulation-based environment. (3) A High Detail Representation-Low Detail Learning Procedure condition in which the learners were provided with exact values for parameter representations, but they were not provided with detailed procedures for learning. (4) A Low Detail Representation-Low Detail Learning Procedure condition in which neither exact parameter values nor learning procedures were provided in detail.

The High Detail Representation-High Detail Learning Procedure condition provided the highest level of instructional guidance whereas the Low Detail Representation-Low Detail Learning Procedure condition provided the lowest level of guidance in the current study, with both the Low Detail Representation-High Detail Learning Procedure condition and the High Detail Representation-Low Detail Learning Procedure condition providing intermediate levels of guidance. Within the theoretical framework of cognitive load theory, an interaction between levels of instructional detail and learner expertise was hypothesized. More specifically, it was expected that the highly detailed learning guidance would be of benefit for learners with little expertise in the domain. With higher expertise, the benefits of more detailed instructional guidance were predicted to disappear or even reverse.

Method

Participants

One hundred and fifty-two (male = 97; female = 55) students participated in the current experiment. All students attended a public high school in northern Taiwan. Twelve (male = 3; female = 9) did not finish the whole experiment and therefore 140 participants were included in the final analyses period. Sixty-nine (male = 46; female = 23) were in grade 10 and had no tuition on the concepts of correlation in school before the experiment except some basic tuition about simple probability and statistical charts such as bar charts, pie charts, etc. These students were classified as lower-expertise, less knowledgeable students in the present study. In contrast, seventy-one participants (male = 44; female = 27) in grade 11 had learned the concepts of correlation in the final term of grade 10, and were classified as higher- expertise, more knowledgeable students. Within both expertise levels the participants were assigned to the four experimental conditions balanced according to their average scores of the previous two midterm mathematics exams.

Experiment environment

The experiment was conducted in the third version of the Simulation-Assistant Learning Statistic environment developed by Liu and his colleagues (Liu et al., 2010) to assist students to learn the basic concepts associated with correlation and revised based on the design of the present experiment and the results of a pilot study conducted one week before this experiment (See Figure 1 for a screenshot of the Simulation-Assistant Learning Statistic III). The Simulation-Assistant Learning Statistic III included two major areas in its interface: a learning guidance area on the left side and a dynamic linked multiple representation area on the right side. Except that the learning guidance varied between conditions, the environment was identical for all learners. The dynamically linked multiple representation area contained a blank for setting up and presenting the correlation coefficient value (r value), a table for displaying and manipulating the two dimensions of each of the sample numbers on the X and Y axes, and a
scatter plot for showing the distribution of data points corresponding to the sample numbers. These three representations presented the concepts associated with correlation in different formats. If the students manipulated any of these representations, the corresponding changes would occur in the other representations simultaneously. All manipulation activities by the students in the current experiment were conducted in this dynamically linked multiple representations area. The learning guidance area included textual instructions varying between conditions. It instructed the learners to set appropriate values for the specific representation and to explore the concepts using appropriate procedures. Specifically, the High Detail Representation-High Detail Learning Procedure and the High Detail Representation-Low Detail Learning Procedure conditions had highly detailed guidance on setting values for specific representations. These values were provided with exact r values (e.g., input 0.8, 1, -0.8, and -1 in this study) and paired x, y values (e.g., five pairs of numbers) while the other conditions with low detailed guidance for the representation settings were only instructed to enter any number between a range of -1 and 1 in the blank for r value and to make any changes for the values of the paired x, y variables. An example of the learning guidance for the four conditions is presented in Appendix A. Similarly, the High Detail Representation-High Detail Learning Procedure and the Low Detail Representation-High Detail Learning Procedure conditions with highly detailed guidance on the way to explore the environment were provided with detailed steps for observing each relation between dynamically linked multiple representations, whereas the other conditions with low detailed guidance on learning procedures were only provided a general direction to explore the relations between dynamically linked multiple representations.

Figure 1. A screenshot of the simulation-assistant learning statistic

Materials and procedure

The experimental materials used in the current experiment were compiled according to the Taiwan National Syllabus Program Requirements designed for Grade 10, Semester 2 mathematics. Three units with two tasks related to three major concepts associated with correlation were selected as the learning materials in the current study, including positive and negative correlations (Unit 1), correlation degree (Unit 2), and perfect correlations (Unit 3). Each unit contained two tasks in which the relations between the dynamically linked multiple representations in the simulation-based learning environment were explored separately by manipulating the r value (Task 1) and the values of the x and y variables (Task 2).
The experimental procedures involved a pre-experimental phase, a learning phase, and a test phase, being executed on two days with an interval of one week. On the first day, only the pre-experimental phase was scheduled, lasting fifteen minutes. On the second day, the learning phase (50 minutes) and the test phase (10 minutes) were administrated using a computer assigned to each participant (see Figure 2, for a schematic representation of the procedure).

Pre-experimental Phase (on the first day)

- Prior knowledge test
  - 8-item multiple choice questions

(After one week)

Learning Phase (on the second day)

Prior instruction and practice before learning

- Condition 1: High Detail Representation-High Detail Learning Procedure
- Condition 2: Low Detail Representation-High Detail Learning Procedure
- Condition 3: High Detail Representation-Low Detail Learning Procedure
- Condition 4: Low Detail Representation-Low Detail Learning Procedure

3 units with 2 tasks learning

(Immediately followed)

Test Phase (on the second day)

- Comprehension test
  - 8-item multiple-choice format questions
  - Cognitive load evaluation

Figure 2. The procedure over the experiment

Pre-experimental phase

A prior knowledge test was carried out one week before conducting the experiment in order to verify the classification of higher- and lower-expertise learners and also to ensure an equivalent initial understanding of the concepts of correlation between experimental conditions within expertise levels. The students were given fifteen minutes to answer eight multiple-choice questions on a sheet of paper. One point was given for a correctly answered question, and full marks for the prior knowledge test were therefore 8 points.

Learning phase

In the learning phase, every student was provided with a computer with the software of Simulation-Assistant Learning Statistic III installed. One page of instruction about basic knowledge of the key concepts associated with correlation (e.g., $r$ value, sample size, category of correlation, etc.) was presented to the students on the computer screen for learning at the beginning of the learning phase. No time limit was required for basic knowledge acquisition, and the students could move to the next page when they finished reading.
After basic knowledge acquisition, the students were introduced to the representations involved in Simulation-Assistant Learning Statistic III as well as their functions by a three-minute video. The video showed how to manipulate the different representations, such as an $r$ value and the values of the $x$, $y$ variables. When the video ended, the students were allowed to practice manipulating the representations on their own for a limit of 1 minute. This time limit was adequate for students to finish their practice. The students were also given opportunities to ask questions after practice.

The students were then presented with the simulation-based learning environment with the learning guidance varying corresponding to their own experimental condition. They were presented with identical learning tasks and an identical simulation-based learning environment except that the amount of detail in the learning guidance varied according to the experimental condition.

**Test phase**

The test phase immediately followed the learning phase and was conducted on the same computer as used in the learning phase. The ten minute comprehension test consisted of 8 multiple-choice items (see Appendix B, for an example question of the comprehension test). The internal consistency of the comprehension tests was 0.61. At the end of the comprehension test, a nine-point Likert rating scale originally designed by Paas, Van Merrienboer and Adam (1994) was presented on the computer screen. Students were asked to indicate their cognitive load in completing the task, with 1-9 being “very very low,” “very low,” “low,” “slightly low,” “neither high nor low,” “slightly high,” “high,” “very high,” and “very very high,” respectively.

**Results**

Statistical significance for all tests was set at the .05 level except when otherwise indicated. Partial $\eta^2$ was used as the effect size index. Accordingly, .01, .06, and .14 are considered as the partial $\eta^2$ values reflecting small, medium and large effect sizes (Cohen, 1988).

**Prior knowledge test**

A 4 (Experimental Condition: High Detail Representation-High Detail Learning Procedure condition, High Detail Representation-Low Detail Learning Procedure condition, Low Detail Representation-High Detail Learning Procedure condition, Low Detail Representation-Low Detail Learning Procedure condition) × 2 (Expertise Level: lower-expertise learners, higher-expertise learners) analysis of variance (ANOVA) was conducted on the prior knowledge test accuracy scores (see Table 1, for the means and standard deviations of the test accuracy scores of the prior knowledge test). The results demonstrated a significant main effect of expertise level, $F(1, 132) = 89.41, MSE = 2.51, p < 0.001$, partial $\eta^2 = .40$, indicating that the division of learners into higher- and lower-expertise levels was valid. Neither a main effect of the experimental conditions, $F(3, 132) = 1.95, MSE = 2.51, p = .13$, nor an interaction between the experimental conditions and expertise levels was found, $F(3, 132) = 1.62, MSE = 2.51, p = .19$, verifying the equivalent initial understanding of the concepts of correlation between the experimental conditions within expertise levels. Based on these results, it was decided not to use the prior knowledge test scores in any other analyses of the present study.

<table>
<thead>
<tr>
<th>Learner expertise</th>
<th>High Detail Representation-High Detail Learning Procedure</th>
<th>Low Detail Representation-High Detail Learning Procedure</th>
<th>High Detail Representation-Low Detail Learning Procedure</th>
<th>Low Detail Representation-Low Detail Learning Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Lower</td>
<td>18</td>
<td>3.61</td>
<td>1.54</td>
<td>18</td>
</tr>
<tr>
<td>Higher</td>
<td>19</td>
<td>5.16</td>
<td>1.64</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Table 1. Means and standard deviations on the prior knowledge test |
Comprehension test

Table 2 shows the means and standard deviations for the four experimental conditions on the comprehension test. We conducted a two-way analysis of covariance (ANCOVA), with four experimental conditions (High Detail Representation-High Detail Learning Procedure condition, High Detail Representation-Low Detail Learning Procedure condition, Low Detail Representation-High Detail Learning Procedure condition, Low Detail Representation-Low Detail Learning Procedure condition) and two expertise levels (lower-expertise learners, higher-expertise learners) as between-subjects factors. The average scores of the two previous midterm examinations in mathematics used to allocate students evenly to the four experimental conditions were used as a covariate.

Table 2. Means and standard deviations for the four experimental conditions on the comprehension test

<table>
<thead>
<tr>
<th>Variables</th>
<th>Conditions</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Detail</td>
<td>Low Detail</td>
<td>High Detail</td>
<td>Low Detail</td>
<td>High Detail</td>
<td>Low Detail</td>
<td>Low Detail</td>
</tr>
<tr>
<td></td>
<td>Representation-High Detail Learning Procedure</td>
<td>Representation-</td>
<td>Representation-</td>
<td>Representation-</td>
<td>Representation-</td>
<td>Representation-</td>
<td>Representation-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low Detail</td>
<td>Low Detail</td>
<td>Low Detail</td>
<td>Low Detail</td>
<td>Low Detail</td>
<td>Low Detail</td>
</tr>
<tr>
<td></td>
<td>Procedure</td>
<td>Procedure</td>
<td>Procedure</td>
<td>Procedure</td>
<td>Procedure</td>
<td>Procedure</td>
<td>Procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-expertise</td>
<td>Performance</td>
<td>18 5.50 1.43</td>
<td>18 4.94 1.63</td>
<td>16 4.56 1.90</td>
<td>17 3.94 1.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cognitive load</td>
<td>18 6.17 1.98</td>
<td>18 6.28 2.24</td>
<td>16 6.50 1.93</td>
<td>17 6.76 1.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher-expertise</td>
<td>Performance</td>
<td>19 5.05 1.72</td>
<td>17 6.24 1.25</td>
<td>18 6.11 0.96</td>
<td>17 6.24 0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cognitive load</td>
<td>19 4.11 1.76</td>
<td>17 3.82 1.98</td>
<td>18 3.72 1.78</td>
<td>17 4.76 1.64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Before analyzing the data using the two-way ANCOVA, the homogeneity of regression coefficients was tested. The results indicated that neither the interaction between the covariate variable and experimental conditions, \( F(3, 130) = .55, p = .65 \), nor the interaction between the covariate variable and expertise levels, \( F(1, 130) = 1.60, p = .21 \), was significant. Thus, neither homogeneity assumptions were violated.

The ANCOVA revealed no main effect of experimental conditions, \( F(3, 131) = .72, MSE = 2.06, p = .54 \). A significant main effect of expertise level was found, \( F(1, 131) = 14.87, MSE = 2.06, p < .001, \) partial \( \eta^2 = .10 \), indicating a better test performance for the higher-expertise learners than for the lower-expertise learners. A significant interaction between expertise level and experimental condition was also obtained, \( F(3, 132) = 5.89, MSE = 2.06, p = .01 \), partial \( \eta^2 = .12 \) (see Fig. 3., for a representation of the significant interaction).

Simple effects tests were used following the significant interaction on the comprehension test performance. For lower-expertise learners, a significant effect of experimental conditions on the comprehension test question accuracy scores was obtained, \( F(3, 131) = 3.52, MSE = 2.06, p = .02, \) partial \( \eta^2 = .08 \). Post hoc LSD tests showed that the High Detail Representation-High Detail Learning Procedure condition and the High Detail Representation-Low Detail Learning Procedure condition significantly outperformed the Low Detail Representation-Low Detail Learning Procedure condition. There were no other significant differences between experimental conditions.

For higher-expertise learners, there was also a significant effect of experimental condition on the comprehension test accuracy scores, \( F(3, 131) = 3.08, MSE = 2.06, p = .03, \) partial \( \eta^2 = .07 \). Post hoc LSD tests indicated the High Detail Representation-High Detail Learning Procedure condition performed significantly worse than any of the other conditions with no other significant differences.

For the cognitive load self-ratings during the comprehension test. The homogeneity of variance assumption was again tested before conducting the ANCOVA. The results indicated neither a significant interaction between the covariate variable and experimental conditions, \( F(3, 130) = 1.22, p = 0.31 \), nor a significant interaction between the covariate variable and expertise level, \( F(1, 130) = 0.84, p = 0.36 \), revealing a plausible assumption of homogeneity of regression coefficients. ANCOVA only revealed a main effect of expertise level, \( F(1, 131) = 33.54, MSE = 3.50, p < .001, \) partial \( \eta^2 = .20 \). Inspection of the means indicated that the lower-expertise learners found the test harder than the higher-expertise learners. The results failed to indicate either a main effect of experimental conditions, \( F(3, 131) = 1.05, MSE = 3.50, p = .37 \), or an interaction between expertise level and experimental condition, \( F(3, 131) = .38, MSE = 3.50, p = .77 \).
Discussion

The aim of this study was to examine the effects of levels of instructional detail in supporting learners in a simulation-based environment. Four varying levels of details in guidance embedded in the same simulation-based learning environment were compared. It was expected that highly detailed guidance would be helpful for lower-expertise learners while lower levels of details in guidance would be beneficial for higher-expertise learners. Overall, the hypothesis was supported by a significant, disordinal interaction of test performance between experimental condition and expertise level on the comprehension test.

A simulation-based learning environment can provide learners with a platform in which learners’ own mental model or schemas can be constructed by independently interacting with the environment. In such a relatively free learning environment, determining what should be dealt with (representation selection) in the environment and how to deal with the selected representation (learning procedure) is critical for effective learning (Kirschner, Sweller, & Clark, 2006). From a cognitive load theory perspective, free exploration of a highly complex environment may result in heavy demands on working memory (Sweller et al., 2011). If minimal details in guidance are provided, learners without sufficient prior knowledge need to devote a large amount of their working memory resources to search for problem solutions, leaving few working memory resources available for learning. The present results supported this hypothesis by demonstrating a positive learning effect using more detailed guidance when learners’ expertise levels were low. Although the High Detail Representation-High Detail Learning Procedure condition did not show a statistically significant superiority on comprehension test performance over the two conditions with intermediate levels of detailed guidance (High Detail Representation-Low Detail Learning Procedure condition & Low Detail Representation-High Detail Learning Procedure condition), the means revealed the expected trend using lower-expertise learners. A significant difference was found between the High Detail Representation-High Detail Learning Procedure condition and the Low Detail Representation-Low Detail Learning Procedure condition. The benefit of more detailed guidance on learning in a simulation-based learning environment was further confirmed by the significantly better comprehension test performance for the High Detail Representation-Low Detail Learning Procedure condition than for the Low Detail Representation-Low Detail Learning Procedure condition.

However, such learning effects due to highly detailed guidance were absent for higher-expertise learners. On the one hand, more knowledgeable learners had sufficient prior knowledge to process multiple elements as a single element when searching and selecting representations, resulting in the cognitive load associated with exploring a complex simulation-based environment being manageable. For example, higher-expertise learners previously had learned the
concepts of correlation and they were likely to know which manipulations would be most beneficial to further advance their understanding. Learning may therefore have been facilitated using a relatively free mode of exploration with few details during guidance. In contrast, highly detailed guidance that provided exact values for setting parameter representations and specific procedures for manipulating the representations already were likely to be part of the prior knowledge base of higher-expertise learners. The cognitive load associated with reading and following guidance was therefore redundant and imposed an extraneous cognitive load for higher-expertise learners, impairing learning.

These results are consistent with previous cognitive load theory studies on worked examples (Kalyuga et al., 2001; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Sweller, Van Merrienboer, & Paas, 1998; Tuovinen & Sweller, 1999) that found that worked-out examples instruction, which was effective for novices, lost its advantage and became redundant as the knowledge level of learners was raised as a consequence of intensive training. For example, Kalyuga et al. (2001) presented either a series of worked examples or a less guided, exploratory-based environment to mechanical trade apprentices. The results showed that when dealing with complex tasks, inexperienced trainees clearly benefited most from the worked examples procedure but the advantage disappeared after two training sessions as participants became more experienced in the domain. Despite the differences in procedure, the current results may be interpreted in the same way as that of the worked example studies: discovery with minimal guidance imposes a heavy cognitive load on less knowledgeable learners, limiting effective learning, especially in complex simulation-based learning environment. The disadvantages of discovery learning disappeared and reversed with higher expertise. These results provide a clear example of the expertise reversal effect (Kalyuga, 2007, 2009a, 2009b; Kalyuga et al., 2003), demonstrating a reversal of cognitive load effects with expertise. The present study, by examining the effects of instructional guidance when exploring a simulation-based environment, demonstrated that whether the instructional guidance imposed a germane or extraneous load on learners was determined by learners’ expertise.

One limitation of the present study lies in the cognitive load evaluations. On the comprehension test, except for the significant difference of the cognitive load self-ratings between higher- and lower-expertise learners, no other effects related to subjective measures of cognitive load were found. The failure to find subjective ratings effects was inconsistent with the test performance where differences between conditions were found. A possible reason for the absent main effects of subjective ratings of cognitive load was that the complex simulation technology imposed some cognitive load by itself and learners may have found it quite hard to differentiate how much cognitive load was imposed by the learning tasks rather than from dealing with the technology. This limitation in cognitive load measurement may be a challenge for all studies investigating technology-assisted learning. Future research on simulation-based learning environment may need to measure the cognitive load by more specific questionnaires that, for example, ask students to rate how much mental effort was spent on the learning tasks separately from how much mental effort was spent on using the technology. Alternatively, repeatedly measuring mental effort after each question of a test may provide a more accurate measurement of cognitive load (Van Gog, Kirschner, Kester, & Paas, 2012).

The relatively low internal consistency for the comprehension test (.61) constituted a second limitation of the present study. The small number of test questions is a possible reason for the reduced Cronbach’s alpha value (Nunnally & Bernstein, 1994). Although the comprehension test questions had been used in previous research with high internal consistency values (Liu, Kinshuk, Lin, & Wang, 2012; Liu et al., 2010), this study only selected eight questions from the previous tests. More test questions should lead to higher internal consistency values.

Another limitation of the present study was its failure to clearly distinguish between the Low Detail Representation-High Detail Learning Procedure condition and the High Detail Representation-Low Detail Learning Procedure condition, which had highly detailed guidance on either representation selection or learning procedures. As a consequence, it is unknown whether representations selection or learning procedure is more important in instructional guidance. Future research may explore this question by further analyzing the cognitive processes during learning, for example, by analyzing the learning performance recorded in the log file or measuring the cognitive load during the learning stage.

In conclusion, individual differences in learner expertise in a domain should be taken into account when selecting user-adapted levels of guidance for simulation-based discovery learning. For novice learners without sufficient prior knowledge, highly detailed guidance should be presented first. After learners become more knowledgeable in the
domain, the guidance can be less detailed in line with the guidance fading effect (Renkl & Atkinson, 2003; Renkl, Atkinson, & Große, 2004; Renkl, Atkinson, & Maier, 2000; Renkl, Atkinson, Maier, & Staley, 2002). These findings may provide a reason for the diverse results in previous research on the effects of guidance for simulation-based learning. For example, van der Meij and de Jong (2011) found a positive effect of learning guidance during computer simulation while Chang, Chen, Lin, and Sung (2008) obtained a negative effect. These apparent contradictions may be explained by the current findings.

Acknowledgments

The authors would like to thank the assistance of Wang-Shen Lu in developing the experiment environment. Moreover, we would like to thank the Ministry of Science and Technology, Taiwan, R.O.C. for financially supporting this research under Grant no. MOST 101-2511-S-003 -061 -MY3, and MOST 103-2511-S-003 -058 -MY3. Finally, we would like to thank the support of the “Aim for the Top University Project” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 104-2911-I-003-301.

References


Appendix A

An example of the learning guidance for the four experimental conditions with different levels of detail

**Unit 1: The concepts of positive and negative correlations**

The default sample size is 20 in this study (n = 20). Please complete tasks 1 & 2 using the following procedures. At the same time, please observe the changes that occur in the r values, paired x and y values, and data points of the scatter plot as well as the relations between them.

**Learn about positive and negative correlations by setting different correlation coefficients (r values)**

<table>
<thead>
<tr>
<th>Condition 1: High Detail Representation-High Detail Learning Procedure</th>
<th>Condition 2: Low Detail Representation-High Detail Learning Procedure</th>
<th>Condition 3: High Detail Representation-Low Detail Learning Procedure</th>
<th>Condition 4: Low Detail Representation-Low Detail Learning Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations Procedure</td>
<td>Given r value</td>
<td>No given r value</td>
<td>Given r value</td>
</tr>
<tr>
<td>Step 1</td>
<td>Set the r value to 0.7</td>
<td>Set the r value to any value ranging from -1 to 1.</td>
<td>Try to observe the changes and the relations between the r values, paired x and y values, and data points of the scatter plot when setting the r value to 0.7, -0.7 as well as pressing the reordering key on the x, y table.</td>
</tr>
<tr>
<td>Step 2</td>
<td>Press the reordering key on the x, y table and observe the relations between the paired x, y values and the r value.</td>
<td>Try to observe the relations between the distribution of data points on the scatter plot and r values.</td>
<td>Once again, reset the r values to 0.9 and -0.9, and observe the changes and relations.</td>
</tr>
<tr>
<td>Step 3</td>
<td>Observe the relations between the distribution of data points on the scatter plot and r values.</td>
<td>Observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
<td>Observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
</tr>
<tr>
<td>Step 4</td>
<td>Observe the relations between the distribution of data points on the scatter plot and paired x, y values.</td>
<td>Observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
<td>Observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
</tr>
<tr>
<td>Step 5</td>
<td>Observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
<td>Observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
<td>Observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
</tr>
<tr>
<td>Step 6</td>
<td>Reset the r value to -0.7 and repeat procedures 2-5.</td>
<td>Reset the r value to its inverse number and repeat procedures 2-5.</td>
<td>Try to observe the changes and the relations between the r values, paired x and y values, and data points of the scatter plot when setting the r value to any pair of inverse numbers ranging from -1 to 1 as well as pressing the reordering key on the x, y table.</td>
</tr>
<tr>
<td>Step 7</td>
<td>According to the same procedure as above, reset the r values to 0.9 and -0.9, and observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
<td>According to the same procedure as above, reset the r values to another pair of inverse numbers ranging from -1 to 1, and observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
<td>According to the same procedure as above, reset the r values to another pair of inverse numbers ranging from -1 to 1, and observe the relations between the r values, paired x and y values, and data points of the scatter plot.</td>
</tr>
</tbody>
</table>

**Task 2: Learn about positive and negative correlations by setting different paired x and y values**

<table>
<thead>
<tr>
<th>Condition 1: High Detail Representation-High Detail Learning Procedure</th>
<th>Condition 2: Low Detail Representation-High Detail Learning Procedure</th>
<th>Condition 3: High Detail Representation-Low Detail Learning Procedure</th>
<th>Condition 4: Low Detail Representation-Low Detail Learning Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations Procedures</td>
<td>Given paired x, y values</td>
<td>No given paired x, y values</td>
<td>Given paired x, y values</td>
</tr>
<tr>
<td>Step 1</td>
<td>Press “clear” to remove all the data.</td>
<td>After pressing “clear” to remove all the data, try to observe the changes and relations between the r values, paired x and y values, and data points of the scatter plot.</td>
<td>After pressing “clear” to remove all the data, try to observe the changes and relations between the r values, paired x and y values, and data points of the scatter plot.</td>
</tr>
<tr>
<td>Step 2</td>
<td>Input six pairs of x, y values that are positively correlated as shown in the following table.</td>
<td>Input any six or more pairs of x, y values that are positively correlated.</td>
<td>Input any six or more pairs of x, y values that are positively correlated.</td>
</tr>
</tbody>
</table>
inputting five pairs of x, y values with a positive correlation and then another five pairs of x, y values with a negative correlation as shown in the following tables.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>40</td>
<td>52</td>
</tr>
<tr>
<td>50</td>
<td>48</td>
</tr>
</tbody>
</table>

inputting any five or more pairs of x, y values with a positive correlation and then repeat with a negative correlation.

<table>
<thead>
<tr>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>

Step 3: Observe the ranking orders of x and y values (whether the x and y values increase or decrease).

Step 4: Observe the changes of r values and think about the relations between the paired x, y values and an r value.

Step 5: Observe the relations between paired x, y values and the distribution characteristics of data points on the scatter plot.

Step 6: Observe the relations between the r value and the distribution characteristics of data points on the scatter plot.

Step 7: Observe the relations between the r values, paired x and y values, and data points of the scatter plot.

Step 8: Press “update” to obtain another six pairs of x, y values with the same r value, and repeat procedures 3-7.

According to the same procedure as above, input six or more new pairs of x, y values with a negative correlation as shown in the following table and observe the changes and relations between the r values, paired x and y values, and data points of the scatter plot.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>68</td>
</tr>
<tr>
<td>30</td>
<td>59</td>
</tr>
<tr>
<td>40</td>
<td>57</td>
</tr>
<tr>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>60</td>
<td>50</td>
</tr>
</tbody>
</table>
Appendix B

An example question from the comprehension test

5. A researcher using a sample size of 9 collected data about how much students’ read literature and their literature test results in order to find the correlation between the two variables. The results indicated a positive correlation but not a perfect positive correlation between the amount of literature reading (X) and the literature test results (Y). Can you please indicate which of the following four sets of data about X and Y represent the data collected in this research?

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>X</td>
<td>23</td>
<td>32</td>
<td>39</td>
<td>43</td>
<td>48</td>
<td>52</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>23</td>
<td>32</td>
<td>39</td>
<td>43</td>
<td>48</td>
<td>52</td>
<td>61</td>
</tr>
<tr>
<td>II</td>
<td>X</td>
<td>31</td>
<td>33</td>
<td>39</td>
<td>41</td>
<td>43</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>42</td>
<td>49</td>
<td>51</td>
<td>53</td>
<td>51</td>
<td>55</td>
<td>52</td>
</tr>
<tr>
<td>III</td>
<td>X</td>
<td>68</td>
<td>69</td>
<td>72</td>
<td>73</td>
<td>81</td>
<td>92</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>71</td>
<td>72</td>
<td>79</td>
<td>76</td>
<td>89</td>
<td>96</td>
<td>94</td>
</tr>
<tr>
<td>IV</td>
<td>X</td>
<td>67</td>
<td>71</td>
<td>74</td>
<td>79</td>
<td>81</td>
<td>84</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>56</td>
<td>61</td>
<td>63</td>
<td>68</td>
<td>77</td>
<td>81</td>
<td>89</td>
</tr>
</tbody>
</table>

A. I, IV
B. I, II, III, IV
C. II, III
D. I
E. IV
Using the Multi-Display Teaching System to Lower Cognitive Load

Tsung-Sheng Cheng*, Yu-Chun Lu and Chu-Sing Yang
Institute of Computer and Communication Engineering, Department of Electrical, Engineering, National Cheng Kung University, Taiwan // tsamcheng@gmail.com // lur.ncku@gmail.com // csyang@ee.ncku.edu.tw
*Corresponding author

ABSTRACT
Multimedia plays a vital role in both learning systems and the actual education process. However, currently used presentation software is often not optimized and generates a great deal of clutter on the screen. Furthermore, there is often insufficient space on a single display, leading to the division of content. These limitations generally increase cognitive load. In this research, a multi-display teaching system was developed, based on the design principles proposed by cognitive load theory (e.g., Sweller, Van Merriënboer, & Paas, 1998) and the cognitive theory of multimedia learning (e.g., Mayer, 2005), to address the need for simultaneous, focused display of multimedia content. The multi-display teaching system was deployed on multiple projectors in real educational environments in order to prove that it is effective in improving learning efficiency. An experiment was carried out with 120 college students as participants. Results showed that multi-display instructional material significantly reduced cognitive load and enhanced learning effectiveness.

Keywords
Multi-display, Multimedia computer assisted learning, Cognitive load, Teaching system, Design principles

Introduction
Teaching materials have diversified from traditional paper and whiteboards into multimedia formats often presented on projection screens. Studies (Lai, 1998; Mayer, 1993, 1997; Mayer & Gallini, 1990; Mayer & Sims, 1994) have shown that integrated multimedia instructional material is beneficial for learning. Similarly, combining text with dynamic multimedia (such as animation or videos) can result in better learning (Channin, 1997; Lai, 2000; Poohkay & Szabo, 1995; Rieber, 1990). To explain findings like these, Daft and Lengel (1984, 1986) proposed information richness theory; which indicates that richer communication media convey information more effectively. Liu, Liao, and Pratt (2009) integrated the technology acceptance model with information richness theory and found that the learner’s degree of concentration increases as presentation formats become more diverse. These new forms of teaching materials, which integrate text, graphics, video, and audio, have been shown to engage the audience much more effectively than traditional means, promoting reading interest and willingness to learn (Vichuda, Ramamurthy, & Haseman, 2001). However, the often limited display space available for projection devices and software that is not user friendly to operate often result in multimedia instructional material being less effective. Instructional materials must often be split into separate pieces in order to fit into available display space. As multi-monitor displays mature, the corresponding software requires prior limits to be transcended and a more flexible environment for multimedia instructional material design to be constructed. This research is aimed to develop a system that supports the presentation and operation of multimedia instructional materials on multiple screens. With properly designed material, the cognitive load of learners can be effectively lowered through the use of this multi-display teaching system.

Theories of cognitive load
There are three main types of cognitive load according to previous theoretical work: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Gerjets, Scheiter, & Cierniak, 2009; Miller 1956, Sweller, Van Merriënboer, & Paas, 1998). Intrinsic cognitive load involves the difficulties inherent in the information itself (stemming from the complexity or difficulty of or the interaction between the material to be learned) and the learners’ degree of expertise (knowledge base or experience) (Sweller et al., 1998). If elements of knowledge are isolated and do not connect with other elements, intrinsic cognitive load will be lower due to this low element interactivity. Conversely, if information is not easily taught on its own and requires complex connections with other elements, learning will require more working memory (a higher intrinsic cognitive load) due to this high element interactivity. Take learning language for instance: intrinsic cognitive load is low when learners simply learn the meanings of vocabulary items that exist in a language (which has low element interactivity). When learning
grammer, in contrast, element interactivity increases, because grammar constitutes the connections among vocabulary items and involves interactions in their meaning. Thus, the same teaching materials may make learners with good related-and-basic knowledge feel that the material is simple and incur little intrinsic cognitive load, while others feel that it is hard and incur more. In this way, intrinsic cognitive load is mainly affected by the connections among elements of the learning objective and by the learner’s intellectual level. Intrinsic cognitive load is hard to alter by means of an instructional design for the reasons that it is mainly caused by learned objects itself (Kalyuga, 2009; Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

Extraneous cognitive load mainly comes from bad material design or low-quality interfaces, which cause learners to consume additional cognitive resources on unrelated information processing during the learning activity (Paas et al., 2003). For instance, some electronic materials are designed with hyperlinks to other websites. Learners who are untrained or inexperienced in computer operation will spend time and effort to learn how to open these links in their browser. Such situations cause learners to need to incur extra effort to learn skills that are irrelevant to the knowledge they are trying to acquire; therefore, extraneous cognitive load is also called “ineffective cognitive load.” This kind of cognitive load can be improved through better design and organization of material.

Germane cognitive load is also called “effective cognitive load.” As noted for extraneous cognitive load above, learners consume cognitive resources during learning activity. Germane cognitive load is caused by material design that assists learners to build a knowledge base including things such as repetition, organization, comparison, and deduction, among others, to help memorization. Germane cognitive load helps with the construction of schemata of learning objects, whereas extraneous cognitive load is in this regard useless. Although germane cognitive load consumes personal cognitive resources, it is effective for learning unless total cognitive load exceeds working memory (Kalyuga, 2009).

Gerjets and Scheiter (2003) determined the relationship among the three types of cognitive load, visualized in Figure 1; it shows that intrinsic cognitive load is mainly affected by complexity of information or knowledge and not improved by teaching style or the design of teaching materials. In contrast, extraneous and germane cognitive load are improved by enhancing teaching methods and materials. Moreover, Paas et al. (2003), Kalyuga (2006), and Van Gog and Paas (2008) have indicated that the extraneous and germane cognitive load can be directly managed by the material designer. An effective instructional design should lower extraneous cognitive load. This accords with the perspective of multimedia learning theory in Mayer (2005) and optimized learning in Van Gog and Paas (2008). How to achieve the purpose of lowering cognitive load and achieving effective learning through a better instructional design has been discussed in the section on instructional design principles below.

![Figure 1. Intrinsic cognitive load, learning complexity, materials, and expertise](image-url)
The measurement of cognitive load

A measurement method is necessary to confirm whether a constructed system can reduce cognitive load caused by teaching material. However, there is no standard measurement of cognitive load. Paas and Van Merriënboer (1994) divided cognitive load into two dimensions: the *task-based dimension (mental load)* and the *learner-based dimension (mental effort)*; both affect learning effectiveness. Wierwille and Eggemeier (1993) proposed three methods for reducing cognitive load: *subjective techniques*, *physiological techniques*, and *task- and performance-based techniques*. Subjective techniques provide learners with a scale for reviewing the learning process. This method, which is the one most commonly used, can quickly record the cognitive process and evaluate it using a single score. Physiological techniques measure subjects’ heartbeats, brainwaves, eye movements, and blood pressure. These physical variations are used to evaluate both cognitive and physical occupational workloads. Task and performance-based techniques measure load based on learners’ performance and the complexity of the tasks involved.

Brunken, Plass, and Leutner (2003) proposed that methods used to measure the cognitive load could be divided by two perspectives: the *objectivity* and *causal relationship* scopes. The scope of objectivity includes methods for subjective self-depiction of information, namely observation of objective behavior, actual conditions, and performance. The causal relationship includes methods of measuring direct value and indirect physiological characteristics (Table 1).

<table>
<thead>
<tr>
<th>Scopes</th>
<th>Causal relationship</th>
<th>Direct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective</td>
<td>Self-reported invested mental effort</td>
<td>Self-reported stress level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self-reported difficulty of materials</td>
</tr>
<tr>
<td>Objective</td>
<td>Physiological measures</td>
<td>Brain activity measures (e.g., fMRI)</td>
</tr>
<tr>
<td></td>
<td>Behavioral measures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning outcome measures</td>
<td>Dual-task performance</td>
</tr>
</tbody>
</table>

Studies from Paas and Van Merriënboer (1994) indicate that subjective techniques are convincible, interference-free, and sensitive to tiny differences. After weighing the costs associated with physiological techniques, we decided to adopt subjective and performance-based techniques to evaluate the effectiveness of the system constructed for this study.

The rating scale technique adopted in this study has been widely used to measure working memory load and mental effort (Paas et al., 2003; Sweller et al., 1998; Van Gog & Paas, 2008), and has been proven to be a reliable measure of both reliability and validity in cognitive load research. This study adopted a subjective cognitive load rating approach modified from previous studies (Cerpa, Chandler, & Sweller, 1996; Kalyuga, Chandler, & Sweller, 2000; Paas, 1992). This scale consists of four questions on a five-point Likert-type scale in two domains, reflecting *degree of clarity* from 1 (very clear) to 5 (very unclear) and *degree of difficulty* from 1 (very easy) to 5 (very difficult). The lowest degree of clarity (the highest score number) represents the heaviest extraneous cognitive load, and the highest degree of difficulty (highest number) represents the heaviest intrinsic cognitive load.

In terms of learning achievement, this experiment evaluated learners by their score on a pretest and a posttest as a transfer assessment designed by the teacher for this experimental course. Transfer assessment was measured by asking students to solve problems using information presented in the instruction. The analysis of results between the pretest and posttest determines the germane cognitive load between the experimental group and the control group.

The design principles of instructional material

Many research efforts have been devoted to finding instructional formats that reduce extraneous load, because it is imposed by processes that do not contribute to learning. While intrinsic cognitive load caused by inherent difficulties and complexities of instructional materials seems unavoidable, extraneous and germane cognitive load can be reduced by improving teaching flow, the teaching methods, the teaching style and the design of instructional materials. To help us understand this situation in more detail, Sweller et al. (1998) proposed several methods that help to improve learning: the *goal-free effect*, the *worked example effect*, the *completion problem effect*, the split-
attention effect, the modality effect, the redundancy effect, and the variability effect. Well-designed materials that utilize these effects to better convey information to learners help reduce cognitive load. Mayer and Clark further proposed multiple principles for multimedia learning (Clark & Mayer, 2011; Mayer, 2002). The success of some of these principles depends on the professional experience of the teachers. For example, in the case of the redundancy effect and Mayer’s coherence principle (put related material closely with less interruption), the teacher decides which content should be included depending on his/her own professional knowledge and teaching experience.

Some principles suggest that display arrangements should involve multiple sensory stimulation. Sweller et al. (1998) and Mayer (2005) both recognized that the simultaneous display of images and sound can heighten learning effects. For instance, Mayer (1997), as well as Tiene (2000) indicated that the modality effect is less easily incurred if these different types of information are displayed sequentially. When instructional knowledge requires learners to reference additional information, make comparisons with others, or integrate some key effects, the simultaneous display of all related information is important to reduce germane cognitive load. Traditionally, limits on display space force teachers to split all related information into chunks. During lectures, teachers demonstrate this information sequentially, thus distracting learners attempting to process each individual piece of information and in turn lowering learning effectiveness. Based on this, teaching assistance systems should support the simultaneous display of multiple multimedia formats. Moreover, Mayer also formulated the signaling principle, which assists with conveying key points from among all display content. The instructor may write down certain points or draw charts to assist learners with organizing information during the lecture. Table 2 summarizes these design principles and the related system requirements.

Table 2. The functionality list according to teaching design principles

<table>
<thead>
<tr>
<th>Sweller’s effects</th>
<th>Mayer’s principle</th>
<th>Cognitive load type</th>
<th>System requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality</td>
<td>Multimedia</td>
<td>Germaine</td>
<td>Display multiple multimedia formats in a single program to lower working memory load</td>
</tr>
<tr>
<td>Split-attention</td>
<td>Spatial contiguity &amp; temporal contiguity</td>
<td>Extraneous</td>
<td>Place course content or information together with commentary and present them at the same time</td>
</tr>
<tr>
<td>Worked example</td>
<td>Signaling</td>
<td>Germaine</td>
<td>Provide colored pens to mark points or provide tips</td>
</tr>
</tbody>
</table>

The limits of existing teaching systems

Following the proliferation of computers, multimedia teaching styles have become popular, because they integrate multiple display methods and make teaching more exciting. However, there are limits to their actual use in practice. The resolution of most projectors is 1920 x 1080, which is not enough to present multiple multimedia content side-by-side. The most common solution is to split all related content into pages, but this degrades communication, for instance by forcing users to separate content and analysis (Kjeldsen, 2006; Lanir, Booth, & Tang, 2008; Parker, 2001; Tufte, 2003), as visualized in Figure 2.

Figure 2. Splitting related content into pages
Splitting content is the least desirable method of displaying it, because it works against instructional design principles that suggest that related information should be displayed closely together. Furthermore, the number and diversity of knowledge formats have increased as information technology has evolved. Teachers inevitably find a need to reference various kinds of files, such as webpages, PDF documents, or videos, to help them explain the material. Studies by cognitive scientists state that spatial and temporal grouping of related items is important for learning (Mayer, 2003): instructors need to allow learners to view the construction of each entity but also to allow them to view entities simultaneously for comparison purposes. Slides do not facilitate this practice easily (Lanir, Booth, & Findlater, 2008).

To address issues like these, Mayer (2005) notes that readers can simultaneously preserve text and images in working memory; and to add some reference points for learners, Erhel and Jamet (2006) suggest the addition of popup windows connected to instructional slides to highlight the focal points of learning. Popup objects can place textual explanations as labels near a corresponding graphical object.

However, a major limitation of popup objects is that they may obstruct the view of texts and images (Chang, Hsu, & Yu, 2011). One may also introduce videos or animations, for instance to introduce real cases related to the material under study. With limited display space, popup windows will make instructional slides appear as shown in Figure 3. Currently, presentation software has little support for features involving cross-type displays; that is, different types of documents cannot be displayed together in the presentation frame. Whenever teachers open referenced material using other programs, the discontinuity and the extra adjustments needed to the display cause extraneous cognitive load. In addition, existing presentation software does not support mouse-track signals across multiple opened windows. To achieve the functionality presented in Table 2, there is thus a need to build a new presentation system.

![Figure 3. Overlapping windows in a single-screen environment](image)

**Multiple-monitor display techniques and the presentation system**

Compared to traditional desktop displays, information presented on multi-display systems is typically separated at a much wider visual angle. Additionally, since displays are often placed at different depths or framed by physical bezels, they introduce physical discontinuities in the presentation of information (Tan & Czerwinski, 2003). Multiple monitors are widely used in the stock market as well as in surveillance systems. This research constitutes a similar attempt to extend display space by connecting multiple projectors in an educational context. A multi-display system benefits teaching efficiency in the classroom because it enlarges the workspace (Seufert, 2003). That is, instructors need much more workspace when presenting multimedia materials contained in several windows, and using multi-display they can place these windows side-by-side on the desktop instead of clicking on the Windows toolbar to constantly switch among windows. A multi-display system contributes more than a single large display because it provides partitions separated by the bezels of the displays, so that instructors can classify their teaching materials naturally by deploying windows showing different media or content to different displays. Students can absorb
different pieces of information efficiently on the basis of this classification by instructors and better connect it with their own knowledge (Pobiner, 2006).

As multiple-monitor setups become more cost-effective and flexible, multiple-screen displays are becoming increasingly common. However, there are only a few presentation software programs available that support expanded display space. Although Microsoft Office PowerPoint, Apache OpenOffice Impress, Apple Keynotes, and cloud-based Acrobat Labs Presentations, Google Docs, and Prezi have developed into mature products, they mostly support no more than two screens. This leads to a situation where one screen, usually a projector, displays the material while another displays the backup notes, intended for the instructor only. Research has revealed that two display screens can not only help improve learners’ recognition memory and peripheral awareness (Robertson et al., 2005), but also result in better task performance and usability (Colvin, Tobler, & Anderson, 2004). Some presentation techniques that have been shown to improve learning effectiveness in this context include reference to previously presented content, visual comparison between different concepts, and non-linear movement through presentation content (Lanir et al., 2008). Hutchings and Stasko (2007) pointed out that another advantage of the multi-screen setup is the ability to support tasks, execute applications, and present images simultaneously.

Many universities and conference centers accommodate lecture halls and conference rooms equipped with multiple high-resolution display systems. Most presentation software does not yet take full advantage of these facilities; in fact, many lecture halls simply default to broadcasting the same slideshow on all of the displays (Lanir et al., 2008). The present research is therefore aimed at the construction of a multi-display system with novel functionality to help teachers adhere more closely and effectively to instructional material design principles. In order to fully utilize the advantages of a multiple-display space, the constructed system includes an editor function with an instinctive interface, meaning one that is easy to operate for the purpose of designing multi-display-fitted materials.

**The multi-display teaching system design**

According to the system requirements discussed in the previous section, the system design details are described as follows:

**Display different types of multimedia contents**

There is no uniform format for multimedia files. For example, video is often used in different formats such as MP4, AVI, and MPEG, among others, not to mention with various encoding techniques. In the system proposed here, a display and viewer object (DVO) is built to handle different format elements. The DVO contains different display components that are responsible for displaying the related file type. Usually the file type is recognized by the file extension name. If the file extension name is unknown, the DVO attempts to display it with a web browser. This also enables the user to configure different display components for the purpose of displaying specific file formats (Figure 4).
Display related multimedia content side by side

To display a larger amount of relevant content, this system utilizes multiple projectors and screens to expand the display space. The system dynamically supports the required number of display screens, depending on the classroom space. Here, we adopted three display screens to demonstrate a usage scenario. The design principles listed in Table 2 were addressed by assigning related content to specific display screens (Figure 5) so that the material could be presented simultaneously. With the display material prepared and configured in this way using the system, instructors can reduce the effort required to open different files in different windows and to arrange the position of each window. Therefore, they can focus more on the content of their lectures rather than the logistics.

![Figure 5. Expanding the display space to display related content](image)

Provide colored pens to mark key points and provide tips

Since the display space is expanded, teachers can better utilize it to make markings on instructional material, thus using the signaling principle to assist students in focusing on key points during the lecture. There is a transparent layer upon which to record pen marks across different screens and even across different multimedia display objects, as shown in Figure 6.

![Figure 6. Transparent layer for drawing lines](image)
Teaching material editor

For better editing and design of material, this system also implements an editing interface. In Figure 7, each row represents a display screen with a sequence from top to bottom. Each column indicates the flow of content to be displayed on each screen simultaneously and in sequence. The operation of this editor is intuitive: users can click on each screen and input the address of a file or webpage. The system supports preview during editing, and users can also drag/import existing PowerPoint files. After the edits are complete, the system will package all files in an FCM file format, portable for use in different places.

![Editing interface of this teaching system](image)

Experiments and evaluation

To verify whether or not the system presented in this research would achieve the expected results, an experiment was carried out using a quasi-experimental method. A subjective approach as well as task- and performance-based techniques were used to evaluate the learners’ cognitive load. We reference the subjective measurement scales proposed by Kalyuga et al. (2000) and Cerpa et al. (1996) and modify them into a five-point Likert-type scale. The scale allows learners to score their cognitive load incurred by learning in the following areas: “The difficulty of the course,” “The effort you made in class,” “The extra effort you made outside of class,” and “The pressure you felt in class.” There are no dependencies among the items in the scale. Learners were asked to recall the learning process they underwent in the course and then answer the questions using the scale.

Participants

The participants in this experiment were 120 college students enrolled in an electronic business class. The content designed under the conditions above and delivered in the class introduces electronic business architecture and related theories, includes specific terms, architectural figures of business models, animations representing the concepts, and
a video of actual cases. This course required cross-reference between multiple files, and as a result was considered suitable as a context for the implementation of this teaching system to teach and to deliver composite knowledge.

**Experimental design**

The research method used is called a “non-equivalent pretest–posttest control group” design. Two groups of students were given a pretest to record their cognitive load level before exposure to the target material. The experimental group was lectured in a classroom equipped with three projectors using the multi-display system (Figure 6). The control group was lectured in a classroom equipped with a single projector. The teachers prepared integrated material edited with the teaching material editor for the experimental group. In contrast, the material prepared for the control group was a PowerPoint presentation with target materials, webpages, and pictures all embedded or integrated in sequential order. The experiment lasted four weeks, three hours every week. The experimental group took class at 8:00 in the morning on Wednesdays and the control group on Thursdays. The course content covered in the first week was an introduction of course guidelines, letting students get familiar with the system. The course proper started in the second week. According to Marcus’s (1996) suggestion, cognitive measurement should be taken immediately after the course to reduce to a minimum the loss of work memory load. Therefore, after every week’s lesson, we used subjective cognitive load rating to measure the cognitive load of students in both groups. In the last week (the fourth week), the posttest was given after the subjective cognitive load rating in order to evaluate the learning effectiveness of the method. The questions in pretest and posttest were selected from the Techficiency Quotient Certification and passed validation by three experts. To control to the degree possible other factors affecting the experiment, the course material and teacher were kept constant in both groups. During the class, the teacher was requested to only use the drawing line function and change page function. Table 3 lists the design and control factors for this experiment.

**Table 3. The experimental design pattern**

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Methods</th>
<th>Scale of cognitive load</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>O1</td>
<td>X1</td>
<td>O2</td>
<td>O3</td>
</tr>
<tr>
<td>Control group</td>
<td>O1</td>
<td>X2</td>
<td>O2</td>
<td>O3</td>
</tr>
</tbody>
</table>

O1: Both groups were given the pretest.
O2: Both groups were given the measurement of cognitive load.
O3: Both groups were given the posttest.
X1: The experimental group was taught using the proposed teaching system.
X2: The control group was taught with a PowerPoint and single-screen setup with various types of multimedia.

**Experimental results**

The four cognitive load scale results from after the four-week teaching experiment are shown in Table 4.

**Table 4. Cognitive load scale results**

<table>
<thead>
<tr>
<th>Group</th>
<th>Items</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>The difficulty of the course</td>
<td>3.03</td>
<td>3.03</td>
<td>3.15</td>
</tr>
<tr>
<td></td>
<td>The effort you made in class</td>
<td>2.93</td>
<td>3.25</td>
<td>3.13</td>
</tr>
<tr>
<td></td>
<td>The extra effort you made outside of class</td>
<td>2.88</td>
<td>3.07</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td>The pressure you felt in class</td>
<td>2.97</td>
<td>3.02</td>
<td>3.05</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>11.85 (3.30)</td>
<td>12.37 (2.73)</td>
<td>12.45 (2.97)</td>
</tr>
<tr>
<td>Control group</td>
<td>The difficulty of the course</td>
<td>3.37</td>
<td>3.38</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td>The effort you made in class</td>
<td>3.37</td>
<td>3.50</td>
<td>3.40</td>
</tr>
<tr>
<td></td>
<td>The extra effort you made outside of class</td>
<td>3.40</td>
<td>3.52</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>The pressure you felt in class</td>
<td>3.25</td>
<td>3.37</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>13.38 (3.30)</td>
<td>13.77 (2.85)</td>
<td>13.63 (3.22)</td>
</tr>
</tbody>
</table>
Next, we analyzed the cognitive load measurement scores with an independent-samples $t$-test. The result is shown as Table 5, showing that the cognitive load of students taught with the MSPS teaching system is lower than that of the control group.

<table>
<thead>
<tr>
<th>Table 5. $t$-test results for cognitive load measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Results of analysis ($\alpha = .05$)</strong></td>
</tr>
</tbody>
</table>

Task- and performance-based techniques can indirectly measure the cognitive load of students by analyzing their learning performance. The descriptive statistics for the pretest and posttest scores are shown in Table 6.

<table>
<thead>
<tr>
<th>Table 6. Learning performance statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Groups</strong></td>
</tr>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>Experimental group</td>
</tr>
<tr>
<td>Control group</td>
</tr>
<tr>
<td>Posttest</td>
</tr>
<tr>
<td>Experimental group</td>
</tr>
<tr>
<td>Control group</td>
</tr>
</tbody>
</table>

Furthermore, we analyzed the pretest and posttest grades with an analysis of covariance (ANCOVA). We exclude the influence of pretest grades (the covariate) on posttest grades (the dependent variable), achieving the standard level of significance ($F = 4.60, p = 0.034$). Also, the grades of the experimental group were higher than those of the control group, showing that the MSPS system certainly helps improve learning effectiveness.

In this research, total cognitive load score were analyzed in relation to the pretest and posttest grades through Pearson’s correlation. Refer to Table 7 below.

<table>
<thead>
<tr>
<th>Table 7. Pearson’s correlation statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Groups</strong></td>
</tr>
<tr>
<td>Experimental Group</td>
</tr>
<tr>
<td>Control Group</td>
</tr>
</tbody>
</table>

Based upon the above results, which indicate an obvious negative correlation, prior knowledge did affect intrinsic cognitive load during the learning process. This result supports the theories of cognitive load presented above.

**Discussion**

In this study total cognitive load was measured as the sum of intrinsic cognitive load and extraneous load (Sweller, 2010). There is no process for separately measuring these three cognitive loads (Paas et al., 2003). In this research, we adopted the same material for both groups in the experiment; therefore, the intrinsic load was assumed to be the same. Since cognitive load as measured by the scale represents total cognitive load, we can deduce that different display modes of material affect extraneous cognitive load directly and analyze the results further to determine if the developed system can reduce the learners’ cognitive load.

Another method to estimate cognitive load during the learning process is through (pretest and posttest) learning performance (Wierwille & Eggemeier, 1993). In this research, the only uncontrollable part was difference in talent/expertise/reational complexity of students. Analysis of covariance was therefore adopted to eliminate these uncontrollable variables. The analyzed results can thus also be deduced to represent the effect of extraneous cognitive load caused by different display methods.

As concluded above, this study demonstrates that the theory Mayer (2005) and Sweller et al. (1998) proposed separately in 2003 and 1998 respectively, claiming that presenting related words and pictures simultaneously and closely is better than presenting them isolated on the other pages, is correct. The multi-display system is able to enlarge the displayable content and make more room for multimedia content in particular to be presented clearly. Although the same effect can be achieved if one has a large enough monitor, high-resolution projectors are expensive, and it costs less to build a 5760 x 1080 pixel presentation space (for instance) by utilizing three common
projectors with resolution of 1920 x 1080 dpi. Moreover, Pobiner (2006) thinks that teachers can classify their teaching material naturally through a multi-displayed system by deploying different media windows to different displays. This can help students absorb different pieces of information and organize this information pre-classified by instructors to form their own knowledge.

Marcus, Cooper, and Sweller (1996) pointed out that in teaching activity, cognitive load is determined by prior experience, nature of the material, and organization. Teachers can only lower the cognitive load through good design and editing of teaching materials and good methods of presentation, though there is not enough evidence to conclude for sure that integrating multiple multi-media benefits the learning process (Bartscha & Cobern, 2003; Mayer, Heiser, & Lonn, 2001; Rieber, 1996; Sweller et al., 1998). Merely adding various media does not make learning more efficient; a teacher’s experience and professional judgment are needed to divide teaching materials into proper lesson-sized portions and to edit them effectively. This study proposes a novel application of a multi-screen system to offer teachers a more flexible tool.

The experimental results show that the multi-display teaching system proposed in this paper does actually decrease cognitive load and efficiently improve learning performance. The system can be extended not only to teaching but also to other tasks such as business presentations. In an interview, the teacher who participated in this experiment pointed out that this approach takes more time and effort than transferring a traditional linear teaching style into a mode divided into previous and next pages or right and left screens. This transfer of material display from a time series dimension to a simultaneous space dimension leads the teacher to need to consider the spatial configuration and choreography more closely. Nevertheless, the teacher saw the learning efficiency of the system as high, and expressed a desire to continue using the system in future teaching.

Due to time limitations, the experiment only lasted for four weeks, with 120 students involved, with a given portion of content (a series of e-commerce course lectures). More precise inferences and larger-scale longer-term experimental realizations need to be carried out in the future to get a more detailed sense of the effects of the system and of potential improvements.

**Conclusion**

Based on design principles for multimedia teaching material elaborated in previous studies, this research built a multi-display teaching system. The system not only allows multiple pages to be presented simultaneously on multiple screens but also enables multiple media to be opened and used in a uniform manner. The system further supports the drawing of functions across screens and media types. Moreover, the teaching system includes an editing interface to improve the design of multi-display instructional material. Ultimately, the purpose of this multi-display teaching system is to lower cognitive load, a benefit that has been verified elementarily in our experiment. This system will be promoted in schools and will be continuously improved as we receive more user feedback.

**References**


An Analytics-Based Approach to Managing Cognitive Load by Using Log Data of Learning Management Systems and Footprints of Social Media

Cheng-Huang Yen1*, I-Chuan Chen2, Su-Chun Lai2-3 and Yea-Ru Chuang4

1Department of Management and Information, National Open University, Taipei, Taiwan // 2Business Administration, Fu Jen Catholic University, Taipei, Taiwan // 3Information Technology Center, National Open University, Taipei, Taiwan // 4Department of Information Management, Fu Jen Catholic University, Taipei, Taiwan // yench@mail.nou.edu.tw // ycchenroc@gmail.com // pure@mail.nou.edu.tw // 013486@mail.fju.edu.tw

*Corresponding author

ABSTRACT
Traces of learning behaviors generally provide insights into learners and the learning processes that they employ. In this article, a learning-analytics-based approach is proposed for managing cognitive load by adjusting the instructional strategies used in online courses. The technology-based learning environment examined in this study involved a video conferencing system and learning management system (LMS) for hosting course content and discussion forums. The social networking software Line was used to enhance the social presence of learners. Students (N = 869) enrolled in a summer course participated in a 9-week experiment. Their LMS log data and social media footprints were recorded, and content experts assessed the intrinsic cognitive load (ICL) of each content module through a consensus process. A learning analytics method was applied to identify candidate parameters relating learning behaviors to cognitive load. The instructor assessed the learners’ cognitive processes and adjusted the instructional strategies according to the results of statistical, discourse, and qualitative analyses. Practical guidelines related to various cognitive load effects were designed to assist the students with managing their cognitive load by using learning behaviors and analytics data as signals for making a change in learning processes. Teachers of online courses can use the proposed approach as a support tool to identify learning problems and assist learners with maintaining a cognitive load that is conducive to learning.

Keywords
Cognitive load, Element interactivity, Cognitive load effects, E-learning analytics

Introduction
The cognitive processing capacity of human memory is limited, as discussed in the literature on limited capacity assumptions (Mayer, 2005). Students participating in online learning experience changes in their cognitive load over time. Instructional activities as well as the topics discussed on forums and the interaction during synchronous video conferences contribute to the cognitive load of online students. Cognitive load varies dynamically with the learning process. Paas and van Merriënboer (1994) proposed an instructional design model for providing instructional strategies for controlling cognitive load during training in complex cognitive tasks. Mental-effort-based measures were recommended as suitable tools for investigating and determining the cognitive load of instructional manipulations. In our research, an analytics-based approach was proposed for measuring and managing cognitive load. Measurements of cognitive load were associated with learning behaviors that were captured by learning management system (LMS) log data and social media footprints.

LMSs enable teachers to monitor online student learning through system-generated reports. Teachers may query the system to determine the amount of time that their students have spent browsing course content or the number of discussion forum posts to which they have contributed. The size and value of the log data from LMSs can be called a type of big data, as discussed by Snijders, Matzat, and Reips (2012). Theories and technologies have been developed to mine actionable information from big data in various application areas. In this research, we explored the possibility of using educational platform log data to manage the cognitive load of learners.

Managing cognitive load by using knowledge of learning behaviors and analytics computing
Cognitive Load Theory (CLT) is a psychological learning theory that has provided a basis for exploring instructional design and learning processes with human cognitive architecture (Sweller, van Merriënboer, & Paas, 1998; Sweller, Ayres, & Kalyuga, 2011). Three sources of cognitive load were identified in CLT: intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL) (Sweller et al., 1998). ICL is determined by the
inherent nature of the learning material and learners’ prior knowledge. ICL cannot be changed by instructional interventions. ECL is determined by instructional design, including how instructional content is designed and the activities required of learners. Sweller et al. (1998) described GCL as the effort involved in relating prior knowledge to current instructional content in order to construct schemas stored in the long-term memory. The three types of cognitive load are additive and contribute to a learner’s global cognitive load that cannot exceed his or her cognitive resources. Otherwise the learning process will fail. On the other hand, learning is ineffective with no or low cognitive load. Although high ICL and ECL is detrimental to the learning process, GCL can help learners to construct schemas that are stored in long-term memory and can be retrieved later in learning or problem-solving. According to CLT, the aim of the instructional design is to reduce ICL and ECL, and to induce GCL while keeping the global cognitive load under the limit of learners’ cognitive resources.

The level of ICL for a learning task is determined by the level of element interactivity and learner’s prior knowledge (Sweller, 2010). An element can be a concept or a procedure that needs to be learned. Low element interactivity materials allow individual elements to be learned with minimal reference to other elements and impose a low working memory load. The concept of element interactivity can be used to explain why some material is difficult to learn or understand (Sweller, 1994).

Figure 1 illustrates the relationships among learning behaviors, log data, and students in a technology-based learning environment. In online courses, students are provided instructional material, a learning schedule, and learning activities. Because all of these tasks are performed online, learner behaviors can be recorded as LMS and social network log data. In this research, we assumed that patterns in learning behavior indicate changes in cognitive load. Paas and van Merriënboer (1994) described the causal factors of cognitive load as characteristics of a task or those of the person performing the task. In our research, the tasks performed by learners include browsing course content, participating in asynchronous discussions and synchronous video conferences, and interacting through social media platforms. Completion of these tasks left traces in system log data, and these traces represent learning behaviors. To assess the cognitive load of learners, learning behaviors must be related to the assessment factors in such measurement dimensions as mental load, mental effort, and performance.

The proposed approach is designed to assist online teachers with alleviating student cognitive load (which depletes mental resources) and enhancing students’ GCL (which leads to effective learning). To achieve this goal, learning problems must be identified through feedback from analytics computing and instructional strategies must be modified to optimize the cognitive load. Technology-based learning environments facilitate computer-supported
collaborative learning (CSCL). Bannert (2002) indicated that controlling cognitive overload is considerably more difficult when several learners are involved in the learning process (e.g., social learning scenarios in CSCL), possibly explaining why CLT-based assumptions are typically explored in individual learning scenarios. Because CSCL plays a critical role in online instruction, we considered scenarios involving multiple learners; which are common in asynchronous discussions and synchronous video conferences. We distinguished two views of cognitive load: micro and macro (Figure 2).

The micro view of cognitive load is focused on the cognitive load of individual learners. When log data are available, online teachers can use them as a guide to provide individualized assistance for learners. For example, if the log data show that a learner repeatedly browses the content of a certain module, then the teacher can attempt to identify the learner’s problem and provide relevant assistance. Once the problem is solved, the learner’s cognitive load regarding the content of the module may decrease sharply and his or her learning behavior may change (e.g., their browsing patterns may return to normal). On the other hand, the micro view of cognitive load can also be used by the learner for self-regulated learning. They may adjust the pace of learning to facilitate an internal management of cognitive load. Because the behavior of every learner is recorded in the system log, the parameters identified using analytics methods reflect learner group behavior. When these parameters correlate with learner cognitive load, they may provide an overall profile that explains the learning process.

The macro view of cognitive load is focused on the cognitive load of individual learners. When log data are available, online teachers can use them as a guide to provide individualized assistance for learners. For example, if the log data show that a learner repeatedly browses the content of a certain module, then the teacher can attempt to identify the learner’s problem and provide relevant assistance. Once the problem is solved, the learner’s cognitive load regarding the content of the module may decrease sharply and his or her learning behavior may change (e.g., their browsing patterns may return to normal). On the other hand, the micro view of cognitive load can also be used by the learner for self-regulated learning. They may adjust the pace of learning to facilitate an internal management of cognitive load. Because the behavior of every learner is recorded in the system log, the parameters identified using analytics methods reflect learner group behavior. When these parameters correlate with learner cognitive load, they may provide an overall profile that explains the learning process.

Figure 2. The difference between a macro view and a micro view of cognitive load

The macro view provides an overall assessment of the cognitive load of a group of students. The teacher can apply appropriate teaching strategies that affect the learning behaviors of the entire class to optimize learners’ cognitive load. In this research, we focused on extracting information from log data to provide a macro view of learner cognitive load. An analytics approach was adopted to transform the log data into meaningful and actionable information.

Timing for assessing cognitive load and for making an instructional intervention

Cognitive load represents the resources used by working memory at a certain point in time. Previous studies have surveyed learners at the end of a learning period to assess their cognitive load for the entire learning period. To identify learning problems in a timely manner, cognitive load must be assessable at any moment in time. Van Gog, Kirschner, Kester, and Paas (2012) found that repeatedly measuring mental effort after performing individual tasks in a series was favored for tasks that take longer than usual to complete. If analytics computing can be used to relate log data with the measurement of cognitive load, then the notion of instantaneous load (Xie & Salvendy, 2000) can be implemented. The changes in temporal patterns of learning behaviors may be explained from the cognitive load perspective.

In this research, the ICL of the learning material was assessed by content experts based on an estimate of element interactivity. Under conditions of low element interactivity, instructional intervention to reduce ECL may have no
appreciable consequences (Sweller, 1994). With the knowledge of the estimate of element interactivity, the timing for making a change in instruction can be adjusted to match the learning period when the level of element interactivity is high or when the temporal patterns of log data suggest a high global cognitive load.

Research goals

The approach proposed in this study provides a real-time analytics-based assessment and a macro view of students’ cognitive load based on LMS log data. The design of the research imposed mental work on students gradually through a pre-planned learning organization. This study was conducted to identify the relationships among the learning behaviors, log data, and cognitive load of students. We also explored how adjusting teaching strategies affected student learning behavior and examined the corresponding changes in the log data and cognitive load. Solving these problems can facilitate controlling and managing student cognitive load at an optimal level throughout the learning period. With the knowledge of the students’ learning behavior and their cognitive load, the instructional techniques based on CLT can be applied to achieve efficiency in learning (Clark, Nguyen, & Sweller, 2006; Sweller, et al., 1998).

Literature review

Management of students’ cognitive load is the key to maintaining effective learning. To manage cognitive load, determining how cognitive load is assessed in CLT is essential. Cognitive load measurements must be correlated with the indicators computed from log data by using an analytics method. The students in this study were encouraged to engage in learning by using social media. The more actively the students participated in learning activities, the more log data were available for the analysis.

Measuring cognitive load with just-enough precision and in due time to spot learning problems

Cognitive load refers to the load imposed on a person’s cognitive systems when he or she is engaged in a particular task (Sweller et al., 1998). To improve the learning process, CLT explains that instructional design should consider human cognitive structure and its constraints (Paas, Renkl, & Sweller, 2003). Cognitive load can be assessed by measuring mental load, mental effort, and performance (Paas & van Merriënsboer, 1994). Wierwille and Eggemeier (1993) proposed three major categories of techniques for measuring mental effort, namely, subjective, physiological, and task- and performance-based indices. Each category incorporates numerous assessment techniques. Two classes of techniques for measuring the effort expenditure have been used in previous research (Paas & van Merriënsboer, 1994): techniques that involve using subjective indices (rating scales), and those that involve using psychophysiological indices (e.g., pupil diameter, heart-rate variability, event-related brain potentials).

The level of ICL for a particular task can be determined by the level of element interactivity and learners’ prior knowledge. Sweller (2010) provided an analysis of element interactivity associated with various cognitive load effects. It was also suggested that ECL and GCL can be defined based on element interactivity. In our research, the ICL was assessed by three content experts according to the element interactivity of the instructional material. We explored how learning behaviors change with the scheduled study of instructional material with varying levels of element interactivity. Pollock, Chandler, and Sweller (2002) developed a two-phase, isolated-interacting elements learning approach. Based on their findings, incorporating all the information elements required for understanding in the instructions may overwhelm a learner’s limited working memory. Information was better learnt through the isolated-interacting elements instructional method for certain groups of learners. Clarke, Ayres, and Sweller (2005) explored the impact of the sequence in which the students learn on their performance. In our research, the course content was uploaded to the LMS in two batches. The students were asked to follow the announced schedule to browse the content. To select an appropriate instructional strategy, the cognitive load of learners must first be estimated. Although the estimate does not need to be precise, it should be readily available. By keeping the instructional activities relevant to the scheduled content, the estimated content ICL can be used as a rough estimate for part of students’ mental load while they were participating in those activities.
Applying CLT-based instructional techniques in technology-based learning environments

CLT-based guidelines were used in the design of instructional material to enhance learning efficiency (Clark et al., 2006). Various instructional techniques also used CLT for instructional design (Sweller, et al., 1998). The availability of log data and the knowledge of learning behaviors allow us to be more confident on when and how to apply the various CLT-based instructional design in technology-based environments. The log data on browsing time provide clues on differential learning times. Darabi and Jin (2013) helped learners to make better use of their cognitive resources by providing example postings and controlling how the posts were displayed on one screen to reduce ECL. The level and quality of interaction in online instruction may correlate with the LMS log data and explain learning efficiency from cognitive load perspectives.

CSCL and collective working memory effect

Kirschner, Paas, and Kirschner (2009) indicated collaborative learning may be favored over individual learning for complex learning tasks due to a larger reservoir of cognitive capacity. Collective working memory effect was interpreted as an advantage of group learning in which information processing can be divided amongst the collective working memories of the group members (Kirschner, Paas, & Kirschner, 2011). Assessing cognitive load can facilitate the online adaptation of learning tasks in computer-based learning environments (Paas et al., 2003). In a collaborative learning environment, the concept of cognitive load needs to be extended to accommodate the situation of multiple learners and group behaviors. Because a major part of learning and instruction in technology-based learning environments is based on CSCL, the notion of collaboration load was explored by Dillenbourg and Betrancourt (2006). The effort by communicating learners to create, share, and interpret an external representation of argumentation increases the cognitive load. Various ways for managing cognitive load in CSCL were discussed in (Van Bruggen, Kirschner, & Jochems, 2002). The impact of collaborative learning in asynchronous discussion groups on cognitive processing was studied in (Schellens, & Valcke, 2005). The types of communications (task-oriented and non-task-oriented) were analyzed to explore learners’ discussion behavior. Darabi, and Jin, (2013) designed strategies based on cognitive load theory to improve the quality of online discussion. Eryilmaz, Alrushiedat, Kasemvilas, Mary and Pol (2009) aimed to develop an understanding of cognitive load as a factor supporting or inhibiting students’ participation in online asynchronous discussions.

Social media and social presence as a catalyst for interaction and participation

Social presence refers to a situation in which people communicate through media and are aware of the existence of other people (Short, Williams, & Christie, 1976). Social presence is a crucial predictor of learner achievement and satisfaction (Kim, 2011). Social media provide an open and cost-free platform for real-time interaction. Previous studies have shown that social media can facilitate improving the quality of interaction among learners (Lin, Lin, & Laffey, 2008; An, Shin, & Lim, 2009) and increasing the level of social presence (Hassanein & Head, 2007; Lin et al., 2008).

In this research, the social networking software Line was used to enhance the social presence of students. Students were invited to join Line groups at the beginning of the summer session. Several course activities were designed to exploit the instant messaging functions of Line. In this research, using social media was anticipated to facilitate learners’ interaction and collaboration.

Learning analytics and learning behaviors

Siemens and Long (2011) defined learning analytics as the measurement, collection, analysis, and reporting of data on learners and the context of their learning, the results of which are used to explain and optimize learning and the environments in which it occurs. The behavior of learners is difficult to predict without using traces of their past behavior, and this is a critical concept in learning analytics (Yen, 2013). The importance of learning analytics lies in the information that can be extracted to facilitate improving the quality of online instruction. Collecting and analyzing learner contexts and learning profiles enables observing and realizing student learning behaviors and assisting teachers with adjusting their teaching strategies in a timely manner.
In online learning environments, interaction traces such as discussion forum posts and social media messages can be gathered. Bruckman (2006) indicated the log file data included the content of discourse that required manual inspection to retrieve valuable information. Clark, Sampson, Weinberger, and Erkens (2007) reviewed different analytics frameworks for assessing dialogic argumentation in online learning environments. The posts by learners in discussion forums were coded and analyzed by analytics methods. Dyckhoff, Zielke, Bültmann, Chatti and Schroeder (2012) implemented a learning analytics toolkit to support the iterative process of improving the effectiveness of their courses and to enhance their students’ performance. Jo, Kim, and Yoon (2014) analyzed the log patterns of adult learners using learning analytics and found that an irregularity of the learning interval was proven to be correlative with and predict learning performance. Gibson, Kitto, and Willis (2014) proposed a cognitive processing framework for learning analytics. The data provided by analytics methods were used to explore the cognitive level of learners. Kalyuga (2007) indicated different types of interactivity provide means for managing various sources of cognitive load. An asynchronous communication environments allow learners to manage rate and amount of information processed at one time.

In this research, we adopted analytics computing to provide a macro view of student cognitive load based on an analysis of log data; this method is an objective approach for assessing cognitive load. However, finding evidence and a theoretical basis for correlating cognitive load with learning behaviors and log data is crucial.

Method

In this research, an LMS was used to host course content that students accessed through the Internet. The LMS also featured asynchronous discussion forums, where an online teacher assigned topics for discussion, and synchronous video conferences were held twice for the entire semester, with each conference lasting 2 hours. The research team was granted administrator privileges to access the log data containing the student learning profiles, including the time that they spent browsing specific content modules, login time, and number of posts that they made. The social networking software Line was used to enhance the social interaction among the students.

Research design

Figure 3 depicts an overview of this research. The four steps in Figure 3 indicate the main focus of this research. At Step 1, students learn through various types of virtual classroom; at Step 2, the learning process leaves traces of learning behaviors in log data that can be computed using e-learning analytics to derive actionable information; at Step 3, the online teachers verify the situation and adjust their teaching strategies on the basis of the derived information; and at Step 4, the adjusted teaching strategies affect the cognitive load of learners. In summary, the online teachers respond to the students’ learning behaviors by adjusting their teaching strategies to optimize the students’ cognitive load. The four-step cycle commenced only when the students were engaged in learning. When students did not participate in the learning environment, social media were used as a convenient tool to redirect them to the learning environment.

The following tasks had to be completed before the concept was implemented. First, we determined the type of useful information that was embedded in the log data and how such information could be transformed into actionable information. Second, we determined the effect of various CLT-based teaching strategies on the students’ cognitive load. Finally, we monitored the students’ learning behaviors to determine whether adjusting the teaching strategies optimized their cognitive load and improved their learning performance.

In determining the relationships among learning behaviors, log data, and cognitive load, we assumed that (a) the students had no prior knowledge, (b) temporal variations in the browsing behaviors of students are related to different types of cognitive load, (c) the level and quality of interaction in the asynchronous discussion forums are related to the students’ cognitive load, and (d) the level of interaction in synchronous video conferences is affected by the various cognitive load effects.
Analysis of participants

The course “Smart Phone Operations and Applications” was offered as an online course in the summer of 2014 by the Department of Management and Information at National Open University, Taiwan. In total, 869 undergraduate students enrolled in the course, which was taught online by using digital content, asynchronous discussion forums, and two synchronous video conferences. Social media were used to enhance the students’ interaction and social presence. Most of the 869 participants were female (64.5%). Most participants were aged 45–54 years (41.5%), followed by 35–44 years (32.3%), and most were adult learners that are graduates of senior high school (49.8%) and college (43.8%). Less than half (40.1%) of the participants had 1–3 years of experience with e-learning. More than half of the participants had become smart phone users before they enrolled the course.

Data collection instruments

The time each student spent browsing the content modules was automatically recorded in the LMS. The browsing time indicator, B-Time, is obtained from dividing the accumulated browsing time by the length of the module measured in the scale of time; e.g., if the accumulated browsing time of a module is 2 hours and the module takes 5 minutes to play to the end, the B-Time for this module would be 24 (module time). For each topic in the discussion forum, the number of views for each post on each day was recorded from the first day of the semester. To facilitate the analysis of the interaction in the asynchronous discussion forum, the density of posts (DP) is defined as the average of the total number of posts on that day, the preceding day, and the subsequent day. For example, if four, six, and five posts were made on 3 consecutive days, then \(DP = (4 + 6 + 5) / 3 = 5\) (for the second day in that series). In related research, DP of a certain day was defined as the number of new posts on that day (Dringus & Ellis, 2010). Since the effect of the instructional events typically lasts for more than one day, we chose the 3-day average number of new posts to define DP. Whether the difference in the definition of DP affects the empirical results will be explored in future research. The increased number of views (IV) is defined as the total number of views for the posts on a certain date.

The messages on Line were saved as text files (one file per Line group). At the end of the semester, the students completed questionnaire surveys, which contained an assessment of ICL for the content of each module (rated using a 10-point difficulty scale), an assessment of ICL for every discussion topic (also rated using a 10-point scale), and instruments for measuring social presence (revised version of the questionnaire by Kim, 2011) and learning satisfaction (4 questions on instructional design and 6 questions on course evaluation). The pretest questionnaire
comprising 20 multiple-choice questions was administered at the beginning of the summer session, and a posttest (also comprising 20 questions) was conducted at the end of the summer session.

**Measurement of intrinsic cognitive load**

The three content experts who taught the course and designed the content were asked to assess the ICL of the content of each module by using a 10-point scale, ranging from 1 (*very low level of element interactivity*) to 10 (*very high level of element interactivity*). The ICL was determined by element interactivity. The information recorded during the assessment process included element interactivity, element complexity, the number of elements, the type of the content (procedures, concepts, facts, processes, or principles), and the nature of the content (technical, theoretical, or practical). The assessment was not precise since we assumed that the students had no prior knowledge. Without the assumption, the estimate of element interactivity is unobtainable. Therefore, the explanation of the results should take this factor into account. The pretest was designed to evenly cover the entire course content. The pretest results were subjected to a Pearson’s correlation test. The error rate correlated positively with the ICLs of the corresponding content modules with significance ($r = .447, p < .05$, two-tailed). In this research, the content modules with high element interactivity were the focus of the study. Under the conditions of high element interactivity, managing the cognitive load is critical to learning efficiency.

The differences in the assessment of ICLs among the content experts were resolved through repeated discussion and consultation with the grading tutors assigned to the course. The pretest results were reviewed along with the assessed ICLs in detail repeatedly. The aim was to ensure the assessment can be a reliable estimate of the mental load imposed by the content. The assessment process was designed to achieve expert validity. The ICL questionnaire was subjected to a reliability test. Because the Cronbach’s $\alpha = .958 > .7$, the ICL rating scale questionnaire achieves high reliability.

**Procedures**

The course was taught without a textbook. Online multimedia content was produced for the course and was organized into ten chapters. The course content was uploaded to the LMS in two batches. The first batch comprised the first six chapters and the second batch comprised the final four chapters. The first and second synchronous video conferences were held in Weeks 4 and 7 of the semester, respectively. Five discussion topics related to the scheduled course content were announced in sequence in the asynchronous discussion forum. The time that each student spent browsing the content modules was recorded in the LMS log data. The number of views for each 24-hour period for every post was recorded at 10:00 A.M. daily. The students were invited to join Line groups that were set up by the teacher. The first synchronous conference was led by the teacher without offering any topics for discussion. Three discussion topics were announced for the second synchronous video conference. Approximately 10 min was allocated to each topic for the students to interact in the Line groups and in the LMS. The leader for each Line group was requested to summarize the discussions in his or her group. The teachers monitored the progress of the discussion in all Line groups and reported to the host of the video conference.

**Adjustment of instructional strategies to manage cognitive load**

Instructional control of cognitive load may be triggered by temporal variations in log data. The instructional strategies can be adjusted with the knowledge of the various cognitive load effects. A summary of how learners’ cognitive load was managed in our experiment is listed in Table 1. Abnormal learning behaviors raised a signal to attract the teacher’s attention. The signal must be verified to remove the possibility of a false alarm (Macfadyen & Dawson, 2010). At the beginning of the semester, the number of logins and B-time indicated whether learners had started to learn. The strategy at this stage was to encourage learners to log into the LMS and browse the course content. The content modules on which the learners focused their mental effort can be identified by the statistics of B-Time. By controlling the accessibility of the course content, the learners were directed to follow the learning schedule. Increased B-time of high element interactivity content may imply high CL. The teacher should review the discourse in the discussion forum and Line to spot learning problems and try to reduce ECL. The design of topics for discussion was a common strategy to direct students’ attention toward effective learning. Low IV and low DP indicated the level of interaction in asynchronous discussion forums was low. The strategy was to encourage students
to participate and interact with peers while ensuring the ECL or collaboration load was kept at a low level when element interactivity was also high. Although increasing the level of interaction may also increase the CL, effective learning induced by CSCL was expected to reduce the CL.

Table 1. The management of learners’ cognitive load (when and how)

<table>
<thead>
<tr>
<th>Learning behaviors reflected in logs or analytics</th>
<th>Possible implication</th>
<th>Suggested actions</th>
<th>Expected effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low number of logins</td>
<td>Possibly no or low CL due to lack of learning activities. Possibly high CL due to learning difficulty. May be a false alarm if other learning data are normal.</td>
<td>Check the students’ learning profiles and verify if they are far behind the learning schedule or have no online activities. Ask students to learn by direct contact or by reminders posted on social media. Help resolve learning difficulty.</td>
<td>The increase in number of logins, the increase in students’ online activities, and variations in CL. Impose a proper amount of stress and CL could promote learners’ cognitive performance.</td>
</tr>
<tr>
<td>Very low B-time</td>
<td>Possibly no or low CL. Possibly high CL due to learning difficulty. May be a false alarm; e.g., learners have related prior knowledge.</td>
<td>Identify the learning problem to decide the proper action. Announce reading assignments and quizzes to impose a proper amount of stress. Help resolve learning difficulty.</td>
<td>The increase in B-time and variations in CL. Learning difficulty resolved.</td>
</tr>
<tr>
<td>Low IV and low DP</td>
<td>Low collaboration load. Possibly no or low CL. Possibly high CL due to learning difficulty.</td>
<td>Encourage participation via intervention or social media. Intervene by providing sample replies or summaries. Help resolve learning difficulty.</td>
<td>The increase in participation, interaction and CL. The improvement in quality of discussion. Learning difficulty resolved and CL reduced.</td>
</tr>
<tr>
<td>Increased B-time</td>
<td>CL may be normal. Possibly high CL due to intensive learning.</td>
<td>Check whether the element interactivity is high. Take action to reduce ICL. Check whether there are sources of ECL. Take action to reduce ECL.</td>
<td>Browsing behavior back to normal. The reduction in ICL and ECL. The increase in GCL.</td>
</tr>
<tr>
<td>High IV and high DP</td>
<td>Non-task-oriented issues that lead to high CL. Task-oriented issues that lead to high ECL and collaboration load.</td>
<td>Resolve non-task-oriented issues; e.g., resolve conflicts. Provide a summary of task-oriented posts to reduce ECL. Spot and resolve learning problems by moderation or intervention.</td>
<td>The improvement in quality of discussion. The reduction in ECL. The increase in GCL.</td>
</tr>
<tr>
<td>High number of task-oriented posts in video conferencing</td>
<td>High ECL and high collaboration load.</td>
<td>Provide feedbacks and summaries.</td>
<td>The reduction in ECL and the increase in GCL.</td>
</tr>
<tr>
<td>High number of non-task-oriented posts in video conferencing</td>
<td>High ECL and high collaboration load.</td>
<td>Direct attention to task-oriented dialogue.</td>
<td>The reduction in ECL and the increase in GCL.</td>
</tr>
</tbody>
</table>

In synchronous video conferences, we focused on inducing students’ GCL. The social networking software Line was used for group discussion. The strategy was to direct students’ attention to task-oriented dialogue and encourage them to participate. Although the use of Line introduced another source of information that may impose high ECL, the learners’ attention was directed to group discussions in Line instead of using the LMS’s text module that was difficult to use and might induce high ECL. The arrangement of learning activities should consider the effect of high
element interactivity; e.g., more instruction time and learning time was allocated to allow the activities of high element interactivity content to be interleaved.

**Results**

Assuming that learners had no prior knowledge, the three content experts assessed the ICL of the content modules based on the level of element interactivity. A questionnaire survey was conducted to determine the ICL of the course content from the students’ perspective (also measured using a 10-point difficulty scale). The results of correlation tests of the cognitive load assessed by the teachers and that assessed by the students (Table 2) indicated that the correlation was not significant ($p = .1$). We speculate that ECL induced from learning material may be a factor for the difference in the assessment of cognitive load by the experts and by the students. Several students complained that the streaming video content for smart phone operations was not easy to follow. They had to stop the video, tried on their smart phone, and started the video again several times. ECL affected students’ evaluation of the difficulty of content. The assessment of ICL by content experts did not consider ECL. Students’ prior knowledge might also be the reason for the inconsistency. The questionnaire survey was conducted after learning. This prevented students from giving a more accurate assessment of difficulty.

<table>
<thead>
<tr>
<th>Cognitive loads by teacher</th>
<th>Cognitive loads by students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Cognitive loads by students</td>
<td></td>
</tr>
</tbody>
</table>

**Browsing time and cognitive load**

Figure 4 shows the curves of the students’ browsing time (B-Time) and the ICL for all content modules (B-time was computed using the average B-time of all students; ICL was assessed by the content experts based on element interactivity). In several cases, the content modules with low ICL coincided with long browsing times (e.g., Module 7-2-1 “Management of Photos on Smart Phones” and Module 8-3-2 “Using Line on Smart Phones”). The figure provides a macro view of cognitive load and learning behaviors because all of the students were considered collectively instead of individually. Although the data shown here were computed at the end of the semester, teachers could retrieve these data anytime by querying the LMS to review the temporal variations in browsing behaviors.

![browsing time vs. ICL](image)

**Figure 4.** The browsing time (B-Time) and the intrinsic cognitive load (ICL)

Pearson’s correlation test was used to identify the relationship between browsing time and ICL for each content module. The results in Table 3 show that the correlation was positive and significant. Both Module 7-2-1 and 8-3-2 contained captured screens of smart phones to illustrate the steps of operations. Several students complained that they had to listen to the instruction carefully because the position of the screen touched by the finger was not shown in the content. This could be a source of ECL. The variations in browsing time were also affected by class events. The peaks of accumulated B-time were observed close to the exam dates and two video conferences. Since ICL was assessed with a general assumption about learners’ prior knowledge, it could only be used as an estimate. On the
other hand, ECL also played a role. Technically, the students might click and play the module and do something else. The interference of these factors makes it hard to use B-time as a valid indicator for cognitive load. In future research, we may consider the learning time (the span of time browsing the content module) and the percentage of completion (for each content module) as candidate indicators.

| Table 3. The correlation matrix of the ICL (assessed by the content experts) and the browsing time |
|---------------------------------------------------|--------------------------------|
| ICL                                               | Browsing time |
| ICL                                               | 1             | Browsing time |
| Browsing time                                     | .265*         | 1             |

Note. *p < .05.

In Figure 5, a scatter chart shows the distribution of ICL versus B-time. The three points in the upper right corner include those content modules with high ICL and high B-time. These points are called hot spots because the corresponding element interactivity assessed by content experts is high. By checking the difficulty level assessed by students, we found that the assessed difficulty levels are also high. These three content modules all include the streaming video with captured screens of smart phones. We speculate that ECL is an issue that affected B-time. Instructional control of cognitive load should be given higher priority to reduce ICL and ECL for these modules first. The points in the lower right corners of Figure 5 correspond to those modules with high ICL and low B-time. The teacher should check and see if there are any learning problems related to these modules. The suggested actions should be taken upon the variations in browsing behaviors. For those modules with low interactivity, we assumed that the global cognitive load would not deplete the students’ cognitive resources. When the log data implies that element interactivity plays the major role, the isolated-interacting elements effect or the molar-modular effect can be used to change the intrinsic nature of what is scheduled to be learned.

Asynchronous discussion and cognitive load

In the asynchronous discussion forums, the number of posts and the number of views of each post was recorded daily from the beginning of the semester. According to the syllabus, 10 chapters were taught in a 9-week summer session. The level of interaction in the discussion forums was measured using an analytics approach (Yen, 2013). The analysis involved determining the density of posts (DP), the increased number of views (IV), and the ICL (element interactivity assessed by the content experts).

Figure 6 shows the recorded data plotted as three curves. The values of the three indicators were normalized to a scale of 100. We sought to determine whether the level of interaction correlated with the corresponding ICL of the content modules for the same period. For example, the content experts rated the ICL of Module 6-2-1 of Chapter 6 with a score of 8 (on a 10-point scale). Chapter 6 was covered in early August, and, accordingly, the DP and IV curves exhibited small peaks during this period.
The course activities typically coincided with an increase in the level of interaction in asynchronous discussion forums. The forum discourse-analysis results indicated that only approximately 30% of the posts were directly related to the content of the course. In other words, we detected the temporal dynamics of the discussion forum but were unable to provide a precise explanation for the changing levels of interaction without performing a manual discourse analysis. We found the increased number of views for most posts reached a peak not long after their appearance in the forum. This implies that most students tended to browse the content of new or unread posts first. From the perspective of cognitive load, the teacher may choose to intervene and reduce the cognitive load if any learning problems were spotted in the new posts. Several students mentioned they were bothered by the increasing number of new posts. They experienced a pressure to read these posts. However, they often found the content of the posts hard to read. They were also concerned about the correctness of the content provided by their peers. When DP is high, it takes time to search for a specific post. The posts in the same thread contain redundant content of previous posts. These scenarios all contributed to ECL.

Pearson’s correlation analysis was used to test the correlations among DP, ICL, and IV (Table 4). The correlation between DP and ICL was negative and significant ($p < .05$); the correlation between DP and IV was significant and positive; and the correlation between ICL and IV was nonsignificant.

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>ICL</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>1</td>
<td>-.261*</td>
<td>1</td>
</tr>
<tr>
<td>ICL</td>
<td>-.261*</td>
<td>1</td>
<td>-.112</td>
</tr>
<tr>
<td>IV</td>
<td>.276*</td>
<td>-.112</td>
<td>1</td>
</tr>
</tbody>
</table>

* $p < .05$.

Krathwohl (2002) discussed six categories of cognitive processes, including remember, understand, apply, analyze, evaluate and create, ordered from low to high levels of cognitive processing. The taxonomy was used as indicators of discussion quality. A total of 179 out of all the 562 posts were identified as related to the course content. Different CLT-based strategies were applied to facilitate the discussion of 5 topics. A sample reply to the topic of discussion was provided for topics 3, 4 and 5, respectively. A summary was provided for each of the topics 2, 3, 4 and 5. The sample reply strategy is analogous to the completion problem effect from the cognitive load perspective. The teacher provided the required format and part of the solution in the sample reply. The summary strategy is similar to the worked example effect in CLT. Some of other students’ replies were included and commented on in the summaries.

The 179 task-oriented messages were reviewed by transcript analysis to classify each message into the categories of low cognitive processing level (i.e., remember and understand) and high cognitive processing level (i.e., apply, analyze, evaluate and create). According to the result shown in Table 5, topics 3 and 4 had the higher number of task-oriented messages. The periods of discussion corresponding to topics 3 and 4 were close to the peaks of the curves of DP and IV. We speculate both strategies had an effect on the level of interaction and the quality of discussion. The result is consistent with the findings reported in (Darabi & Jin, 2013). The period of discussion of topic 5 was close to the end of the semester. The level of interaction was low according to Figure 6.
Table 5. Number of messages that belong to high and low level cognitive processing

<table>
<thead>
<tr>
<th>Topic</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>High level of cognitive processing</td>
<td>3</td>
<td>6</td>
<td>35</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>Low level of cognitive processing</td>
<td>4</td>
<td>10</td>
<td>36</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>Provide sample reply</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide summary</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The IVs for the sample replies and summaries were much higher than that of other posts. Several students mentioned the summaries saved them a big hassle to search through the forum for related posts. This strategy may reduce ECL. Both the summary and sample reply strategies have the effect of keeping the discussion focused on the designated topic. There were several cases where the students were diverted to non-task-oriented discussions. This might increase DP and IV while deteriorating the quality of discussion. Keeping the students focused on the topic of the discussion avoids the split-attention effect from the cognitive load perspectives.

Synchronous video conferences and cognitive load

A total of 112 students attended the first synchronous video conference on July 26, and 158 students attended the second synchronous video conference on August 16. Both conferences ran from 10:00 A.M. until 12:00 P.M. The Line groups were created on August 1. During the second synchronous conference, the discussion topics were announced in the second hour. The number of Line messages increased significantly after 11:00 A.M., and the number of messages related to the topics instead of social chat was higher in the second hour than in the first hour.

Table 6 lists the number of messages related to the assigned discussion topics. Because the three topics were announced in sequence, the correlations between the number of Line messages and the ICL of each topic (assessed by the content experts) were tested, and no significant correlation was observed. After the first topic was announced, several students expressed uncertainty regarding what to expect and when the activity ended. This might induced ECL since the students would try to find out what was going on. The teacher had to lead the discussion and coordinate the activity. The text appeared on the video conference screen along with the teacher’s explanation was a practice of the modality effect to expand working memory capacity.

Table 6. Number of messages related to assigned discussion topics (23 students showed up in the five Line groups)

<table>
<thead>
<tr>
<th>Sources</th>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line group 1</td>
<td>11</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Line group 2</td>
<td>25</td>
<td>39</td>
<td>28</td>
</tr>
<tr>
<td>Line group 3</td>
<td>11</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>Line group 4</td>
<td>20</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Line group 5</td>
<td>8</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Sum</td>
<td>75</td>
<td>101</td>
<td>62</td>
</tr>
<tr>
<td>Synchronous video conference software</td>
<td>1</td>
<td>48</td>
<td>10</td>
</tr>
<tr>
<td>Mean</td>
<td>12.7</td>
<td>24.8</td>
<td>12.0</td>
</tr>
<tr>
<td>SD</td>
<td>3.5</td>
<td>6.32</td>
<td>3.26</td>
</tr>
<tr>
<td>Difficulty scale of topic</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

The second topic was announced and the starting and ending times of the group discussion were stated explicitly. Additional time was allocated for discussion, and the teacher refrained from speaking during the discussion to avoid split-attention effect. The teacher summarized the feedback from each group after every discussion. The pace of the synchronous conference might easily overload the flow of information, contributing to a high cognitive load. The discussion of the third topic indicated that the students interacted effectively. Although the topic covered highly complex sociocultural topics (specifically, the abuse of smart phones), the students addressed the key points of the topic early in the discussion. We speculate that most students became familiar with the process and ECL was reduced. The change in CLT-based instructional strategy facilitated reducing the cognitive load.
A total of 50 students began interacting on Line on August 1. The data in Table 7 indicate that the students’ posts on Line were primarily social for the period of August 1 to August 15. The discussion activities on August 16 caused the students to focus on task-oriented interaction. Few of the students who did not use Line contributed to the interaction by using the default LMS module for inputting text.

<table>
<thead>
<tr>
<th>Date</th>
<th>8/1~8/15</th>
<th>8/16 (7/26 for video conf. 1)</th>
<th>8/17~8/31</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Task-oriented</td>
<td>Non-task-oriented</td>
<td>Task-oriented</td>
</tr>
<tr>
<td>Line group 1</td>
<td>12</td>
<td>65</td>
<td>31</td>
</tr>
<tr>
<td>Line group 2</td>
<td>6</td>
<td>75</td>
<td>92</td>
</tr>
<tr>
<td>Line group 3</td>
<td>8</td>
<td>46</td>
<td>45</td>
</tr>
<tr>
<td>Line group 4</td>
<td>14</td>
<td>42</td>
<td>37</td>
</tr>
<tr>
<td>Line group 5</td>
<td>3</td>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>Sum</td>
<td>43</td>
<td>250</td>
<td>238</td>
</tr>
<tr>
<td>Video conf. 1</td>
<td>N/A</td>
<td>N/A</td>
<td>23</td>
</tr>
<tr>
<td>Video conf. 2</td>
<td>N/A</td>
<td>N/A</td>
<td>59</td>
</tr>
</tbody>
</table>

The data and our observations indicated that students were much more active on social media. If all the conversations were to appear in the text area of the synchronous video conference software, it would be difficult for the participants to work in groups and trace the interaction. Both the teaching staff and students mentioned that scrolling up and down the text area frustrated them. The interface is inadequate and may impose ECL. One of the convenient features of video conferencing software is the real-time poll that can be used to check the students’ learning performance. Although Line messages provided abundant feedbacks from the students, most of these messages were short. Very few of these messages were of high cognitive processing levels based on the taxonomy introduced in (Krathwohl, 2002).

**Learners’ performance**

A paired-samples test was used to compare the students’ performance in the pretest and posttest. Cohen’s $d$ is listed in Table 8 to show the effect size of the experiment. The level of practical significance was determined according to the value of coefficient $d$ ($.2 =$ small, $.5 =$ medium, and $.8 =$ large). The students’ performance in the posttest was significantly higher than that in the pretest ($t = 14.516$, Cohen’s $d > .8$). Future research is planned to include a control group and an experimental group to explore whether managing the cognitive load exerted a positive effect on the students’ performance.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>$t$ value</th>
<th>Cohen’s $d$</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>158</td>
<td>76.519</td>
<td>13.221</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final exam</td>
<td>158</td>
<td>87.505</td>
<td>9.539</td>
<td>14.516***</td>
<td>.953</td>
<td>Large</td>
</tr>
</tbody>
</table>

Note. ***$p < .001$

**Discussion and conclusion**

The assessment of ICL depends on the element interactivity of the course content. Course designers or content experts can only provide an estimate of element interactivity by assuming that the students had no prior knowledge. Moreover, the information is static, unlike the students’ learning behaviors, which can change over time. Because the students learned according to the schedule arranged by the teacher, the static values of the estimated ICL for each content module can be used as a reference for the level of element interactivity of the content learned by the students throughout the semester.
Applying the framework in online instruction

Figure 7 depicts the five-step cycle of the proposed framework. The students learned through online platforms. Log data were generated for the analysis by using an analytics approach. The system raised a signal for action when certain conditions were met. The teacher needed to verify the situation by reviewing the overall picture of the students’ learning process. There might be false alarms. Actionable information was then provided for the teacher to adjust the adopted teaching strategies. The adjustment affected the students’ learning behavior and cognitive load, which was subsequently reflected in the log data. In this research, empirical data were collected from multiple sources, including the LMS log data, survey data, and the data from interaction on Line. Discussion topics for the asynchronous and synchronous interactions and various online activities were designed based on the teaching strategies and according to various cognitive load effects. Occasionally, the teacher provided solutions for pop quizzes on the LMS and posted a message on Line to notify the students. A sharp increase was observed in the level of interaction, first on Line and then on the LMS.

The many facets of learning behaviors

The temporal dynamics of learning behaviors are difficult to capture because multiple factors can contribute to learning behavior. For example, students may spend additional time completing content modules that are difficult, although they may also spend additional time working on easy modules that satisfy their needs in certain applications. Some learners may spend less time on difficult content modules because they are stuck or have prior knowledge. The same situation applies to the asynchronous discussions. Through a manual discourse analysis, we found that some class events had a particularly strong effect on the level of interaction in the discussion forums; for example, many posts appeared on the LMS when the teacher announced the allocation of students to the Line groups.

Keep learning focused and on schedule

Regarding the content modules of high level element interactivity, specific discussion topics were designed and posted in the forum to elaborate on certain elements or to summarize key concepts in order to reduce the material’s cognitive load. Most of the teacher interventions were intended to direct students to focus on the scheduled course content and assigned topics. If the instructions were not kept focused and the course was not conducted on schedule, then it would be difficult to capture a precise indication of student learning behaviors and cognitive load. However, the indicators obtained through the analytics approach are simply numbers for use as a quick reference; thus, they cannot be fully relied upon as indicators. To more accurately monitor the students’ browsing behavior, we recommend releasing the content according to a preset schedule. For example, suppose that the lesson scheduled for Chapter 2 is held in Week 3; the content for Chapter 2 should then be made accessible during Week 1 and closed before the end of Week 3. During the scheduled learning period, the teacher may announce related discussion topics to direct the students’ attention to the content covered in the schedule. Eryilmaz et al. (2009) found that anchored discussions assisted students with engaging in more effective interaction. When the generated log data corresponded
to the specific content modules listed on the schedule, the ICL determined by the content experts would be a valuable indicator of students’ cognitive load.

**Internal management of cognitive load**

Bannert (2002) described the external management of cognitive load as an approach that can be implemented by optimal instruction. On the other hand, the internal management of cognitive load is controlled by learners with their metacognitive and self-regulative competence. Although our research was focused on the external management of cognitive load, the approach can be used to help learners perform their internal management of cognitive load.

**Social media as a tool for attracting more log data and reducing cognitive load**

We found that the students enjoyed interacting on social media. Moreover, keeping the students informed of class events and activities through social media was convenient and effective for the teacher. Most students mentioned that they used social network sites considerably more frequently than they did the LMS. When the online students were involved in a technology-based learning environment, it was observed that using social media allowed them to receive responses or assistance in a timely manner. Furthermore, adopting social media in the course activities improved the students’ social presence and learning satisfaction.

**Limitations of the study**

Because of the large amount of log data, computer programs are required to automate the processing of empirical data. The use of social media has been a challenge for teaching staff, particularly in synchronous video conferences. The students in every Line group became active when the discussion topic was posted. The online teacher required assistance from other staff with organizing and summarizing the interaction so that he or she could provide feedback for the students. However, joining the Line groups was not compulsory for the students, and not participating in the synchronous video conferences did not affect their academic record. The data were collected only for a segment of the student body. The novelty effect of using Line in online instruction should be considered a factor that may affect students’ learning behaviors.

There were technical issues that need attention; e.g., the students may stay on the content page while doing something else. This might raise a false alarm. Not all types of cognitive load left traces in log data. Several students complained about the time that they wasted searching numerous posts to locate a certain post. They also experienced difficulty in keeping track of posts that had not been read. These problems increased the cognitive load.

**References**


Interactivity of Question Prompts and Feedback on Secondary Students’ Science Knowledge Acquisition and Cognitive Load

Kun Huang1, Ching-Huei Chen2*, Wen-Shiuan Wu2 and Wei-Yu Chen2

1Department of Instructional Systems and Workforce Development, Mississippi State University, MS, USA // 2Department of Industrial Education and Technology, National Changhua University of Education, Changhua, Taiwan // khuang@colled.msstate.edu // chhchen@cc.ncue.edu.tw // w.wshiuan@gmail.com // s01261120@gmail.com

*Corresponding author

ABSTRACT

This study investigated how question prompts and feedback influenced knowledge acquisition and cognitive load when learning Newtonian mechanics within a web-based multimedia module. Participants were one hundred eighteen 9th grade students who were randomly assigned to one of four experimental conditions, forming a 2 x 2 factorial design with the presence or absence of question prompts as one factor and types of instructional feedback as the other. With regard to knowledge acquisition, the findings revealed a significant main effect of question prompts and a significant interaction between question prompts and feedback. With regard to cognitive load, the results found a significant interaction between question prompts and feedback. Students who received problem-solving question prompts and corrective feedback achieved better performance and perceived less cognitive load. Implications for designing web-based science learning are discussed.

Keywords

Question prompts, Scaffolding, Problem solving, Feedback, Science learning, Cognitive load, Web-based learning

Introduction

Problem solving is an essential 21st century skill (Trilling & Fadel, 2009), and is continuously incorporated as an integral part of school curricula (Barron, 2000; Barrows & Tamblyn, 1980; Qin, Johnson, & Johnson, 1995). As web-based learning becomes the mainstream in many educational settings, it is increasingly important to adopt research-based guidelines to design effective web-based instruction for problem solving. Two research foci have become prominent: scaffolding and cognitive load. Scaffolding research focuses on designing tools and strategies that provide learners with optimal support as they work on learning tasks. A large body of research investigated the scaffolding of open-ended or ill-structured problem solving (e.g., Chen, 2010; Davis, 2000). Closely related to scaffolding is the provision of feedback to students’ performance as a way to support learning (e.g., Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Corbalan, Kester, & van Merriënboer, 2009).

Cognitive load theories approach the design of web-based learning from a different angle. Focusing on making optimal use of humans’ limited working memory, cognitive load theorists have identified design principles and guidelines that minimize extraneous cognitive load while focusing learners’ cognitive resources on tasks directly related to learning (Sweller, 2010; Sweller, Ayres & Kalyuga, 2011).

While the two lines of research relate to each other, they do not intersect much in the research literature. Presumably, scaffolding which is intended to support problem solving should help to streamline learners’ cognitive processes and facilitate schema construction. Yet existing research appears to cast some doubts on such a presumption (e.g., Chen, Wu, & Jen, 2013; Ge, Chen, & Davis, 2005; Hwang, Kuo, Chen, & Ho, 2014). Therefore, this study set out to implement scaffolding and feedback in a web-based learning environment to support students’ problem solving in science, for the purpose of examining how scaffolding and feedback impact learners’ cognitive load as well as knowledge acquisition.

Theoretical framework

Scaffolding problem solving in science

Solving problems is an essential practice in the disciplines of science. Polya (1957) proposed an influential model to characterize a four-step process in problem solving: understanding the problem, planning a solution, executing the
plan, and checking the result. In the context of solving a science problem, the process involves identifying relevant information in the problem, determining known and unknown concepts, selecting rules or principles applicable to the problem, applying rules or principles, and ensuring that a satisfactory solution is reached (Jonassen, 2000; Simon, 1978). In other words, learners are in on the active construction, manipulation, and testing of mental models of the problem (Jonassen, 2011).

While problem-solving processes are known to researchers and intuitive to skillful problem solvers, students are often not strategic in their problem-solving approaches. Instead of taking time to comprehend a problem and build a conceptual model of it, learners often jump quickly to solutions (Bransford, Brown, & Cocking, 1999). As such, researchers have employed various strategies to scaffold students through the problem-solving process (Arnaud, Arevalillo-Herráiz, Puig, & González-Calero, 2013; Fund, 2007; Palinscar, 1986; Rosenshine & Meister, 1992; Rosenshine, Meister, & Chapman, 1996; van Merriënboer, Kirschner, & Kester, 2003). Question prompting is a frequently used approach to scaffolding learners’ problem solving (e.g., Chen, 2010; Ge, Chen, & Davis, 2005; Saye & Brush, 2002). By presenting questions to students, question prompts focus students’ attention on relevant aspects of problem solving and guide them through the process. Numerous studies have found the effectiveness of question prompts in promoting problem solving, knowledge acquisition, and metacognition (Davis, 2000; Rae, Schellens, Wever, & Vanderhoven, 2012; Zydney, 2010). For example, Ge and Land (2003, 2004) used question prompts to support ill-structured problem solving, and found that the prompts led to better performance in all the major problem-solving processes. In science education, a recent systematic review has identified question prompts to be the most effective scaffolds to regulate student cognition (Devolder, van Braak, & Tondeur, 2012).

While previous studies have mostly focused on the use of question prompts to support ill-structured problem solving, few have examined their use to help students solve conceptual, application problems in science, in which students have to identify the relationships among variables and apply relevant rules and principles to solve them. Research in science education has long established that while students are good at plugging in numbers and using formulas to solve a problem, their conceptual understanding is often lacking and plagued with misconceptions (Hestenes, Wells, & Swackhamer, 1992; Vosniadou & Brewer, 1992). Therefore, in this study, we investigated whether question prompts could help improve students’ performance in solving application problems in the domain of physics.

Because application problems in science are often well structured, having convergent answers and a preferred or prescribed solution process (Jonassen, 2011), question prompts that guide students’ problem solution can take the form of multiple-choice questions, instead of the open-ended question format often seen in studies. By drawing students’ attention to relevant information in a problem, guiding them to understand the problem, and providing them with multiple plausible answers along the process, multiple-choice question prompts offer several advantages. First, they force students to engage in the thinking process and make a specific choice for a given question. Further, the multiple-choice format makes it possible to provide targeted feedback to learners based on their answers, which can be beneficial to web-based learning. In the next section, we review the use of feedback in problem solving.

Feedback during problem solving

Feedback is a critical part of learning. Effective feedback not only helps learners to understand the subject being studied but also gives them clear guidance on how to improve learning. Studies have shown feedback to be strongly and consistently related to achievement regardless of grade, socioeconomic status, race, or school settings (e.g., Corbalán, Kester, & van Merriënboer, 2009; Lee, Lim, & Grabowski, 2010; van der Kleij, Eggen, Timmers, & Veldkamp, 2012). Feedback can be provided immediately upon students’ completion of certain tasks, or after a period of time. Immediate feedback appears to positively influence learning (Arnaud et al., 2013; Corbalán, Paas, & Cuypers, 2010). In teaching science problem solving, immediate feedback was found to be an effective way to correct students’ misconceptions or incorrect connections (Kornell, Hays, & Bjork, 2009; Taconis, Ferguson-Hessler, & Broekkamp, 2001; Vaughn & Rawson, 2012).

Two types of feedback are often provided to students in web-based learning: knowledge of results (KR) and knowledge of correct response (KCR). After learners respond to a question, KR simply informs them whether or not the response is correct; for example, “Your answer is correct.” (Dempsey, Driscoll, & Swindell, 1993). In comparison, KCR provides the same information as KR about the correctness of a response, but additionally specifies the correct answer; for example, “Your answer is incorrect. The correct answer is increase.” (Jaehnig &
Miller, 2007). While KR has an impact on learning (Wulf & Shea, 2004), Lantz and Stawiski (2014) observed that KCR improved students’ ability to retain information and led to deep cognitive processing. In addition, KCR was found to accelerate learners’ conscious construction of schema and reduce random problem-solving attempts (Mea\l or & Dienes, 2013; Scott & Dienes, 2008). From the students’ perspective, they perceived great utility in immediate KCR, and exhibited positive attitudes toward it (Timmers & Veldkamp, 2011; van der Kleij et al., 2012).

Studies comparing different types of feedback generally agree that KCR produced more positive learning outcomes than KR (e.g., Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Jaehning & Miller, 2007; van der Kleij, Eggen, Timmers, & Veldkamp, 2012). However, in past studies, feedback was often provided after students had performed certain problem-solving tasks. Few have examined the use of feedback in conjunction with question prompts that are intended to help students understand a problem before attempting to solve it. That is, feedback is not only provided after students solve a problem, but also during the mental thinking and reasoning processes that lead to problem solution. In such a context, would the two types of feedback, KR and KCR, achieve similar effects as they do in other studies? Further, would question prompts affect science problem solving differently, depending on the type of feedback students receive? Conversely, would the effect of feedback vary depending upon the presence or absence of question prompts scaffolding? The current study intended to address these questions.

Cognitive load

Cognitive load theory views learning as the construction and automation of schemas based on limited working memory capacity (Sweller, 1994). The research literature has examined how different instructional activities influence students’ cognitive load, particularly within technology-enhanced learning environments (Boucheix & Guignard, 2005; Cierniak, Scheiter, & Gerjets, 2009; Schnotz & Kürschner, 2007; Serge, Priest, Durlach, & Johnson, 2013). Three types of cognitive load have been identified: (a) intrinsic cognitive load, which has to do with the complexity of the learning material, (b) germane cognitive load, which are cognitive resources directly contributed to learning, and (c) extraneous cognitive load, which refers to information or activities that are irrelevant to the learning task (Paas, Renkl, & Sweller, 2004). To achieve successful construction of schemas, extraneous load should be kept at a minimum while most cognitive resources should be allocated to germane cognitive activities (Chandler & Sweller, 1991; Paas & van Gog, 2006).

In web-based learning, when novice learners attempt to solve problems without guidance or instruction, they may find the task overwhelming. Supposedly, both scaffolding and feedback provide a way to reduce cognitive load in problem solving, because both mechanisms are intended to provide support so that learners’ working memory is not overwhelmed by the demand of problem solving. Indeed, scaffolding has been argued to reduce cognitive load in problem-based and inquiry learning (e.g., Hmelo-Silver, Duncan, & Chinn, 2007). Yet cognitive load theory casts doubts on the universally positive effect of scaffolding. For example, the expertise reversal effect argues that the effects of instructional guidance vary according to learners’ expertise level. As expertise increases, too much support may actually have a reversal effect on learning (Kalyuga, 2007; Kalyuga, Ayres, Chandler, & Sweller, 2003). Within the empirical literature, few studies have investigated how scaffolding, particularly question prompts, impacts cognitive load in problem solving. Some studies, while not directly measuring cognitive load, appear to suggest that scaffolding can be intrusive and may ultimately obstruct learning (Adams & Clark, 2014; Chen, Wu, & Jen, 2013; Ge et al., 2005). Further, a recent study by Hwang and colleagues (2014) found that a concept-mapping task intended to scaffold web-based problem solving produced high cognitive load within students. Thus, a clear conclusion about the effect of scaffolding on cognitive load cannot be drawn from the existing studies.

On the other hand, research appears to have produced more unequivocal findings about the effect of feedback on cognitive load. For instance, Yeh et al. (2012) found that feedback reduced 10th grade students’ cognitive load when they learned science from animation-based instruction. Moreno (2004) confirmed that explanatory feedback reduced cognitive load in inquiry learning. Yet to the authors’ knowledge, no study has compared the effects of KR and KCR on cognitive load in the context of science problem solving. Would KCR produce less cognitive load than KR in such a context? Would the combination of KCR and question prompts provide optimal problem-solving support, thereby producing minimal cognitive load? These questions remain to be answered.
Research questions

Built on the existing literature, this study examined the use of question prompts and feedback to support science problem solving in web-based learning. The study intended to answer the following questions:

- How does the presence of problem-solving question prompts affect students’ knowledge acquisition and cognitive load?
- How does the type of instructional feedback affect students’ knowledge acquisition and cognitive load?
- How do problem-solving question prompts interact with the type of feedback to impact students’ knowledge acquisition and cognitive load?

Methodology

Design and participants

This study employed a pretest-posttest, $2 \times 2$ factorial research design. The first factor was the presence or absence of problem-solving question prompts (PS or noPS), and the second factor was the type of feedback provided to learners (KR or KCR). One hundred eighteen 9th grade students from four classes were recruited from middle schools in central Taiwan (57 females and 61 males; 14 to 15 years of age). The four intact classes were each randomly assigned to one of the four treatment groups: (a) PS/KR ($n = 27$), (b) noPS/KR ($n = 30$), (c) PS/KCR ($n = 30$), or (d) noPS/KCR ($n = 31$). The four classes were equivalent in student characteristics and academic ability.

Materials

The e-learning environment for this study was a web application written in the Hypertext Preprocessor (PHP) web-scripting language. A database connected to the e-learning environment hosted all the materials used in the study, including instructional presentations, problem-solving tasks and interventions, a knowledge acquisition test, and a cognitive load measure.

Web-based instructional presentations

The instructional presentations consisted of five modules: (a) force and acceleration, (b) the definition of Newton’s Second Law of Motion, (c) the unit and solution strategies in Newton’s Second Law of Motion, (d) the graphics in Newton’s Second Law of Motion, and (e) application examples of Newton’s Second Law of Motion. As shown in Figures 1 and 2, the five modules were represented by the five buttons at the top of the web interface. Each module addressed specific learning objectives.

For example, when we shop in a grocery store, we use a shopping cart …

Figure 1. A sample of web-based instructional presentation
For example, upon completing Module 3, students were expected to: (a) indicate that when the overall external force acting on an object is not zero, the object must have an acceleration, (b) understand that when the mass is fixed, the greater the acceleration, the greater the force, (c) understand that when the force is fixed, the greater the mass, the smaller the acceleration, and (d) explain Newton’s Second Law of Motion and apply the formula \( F = ma \) to solve problems. All the modules were produced using Adobe Flash CS3 which incorporated multimedia elements such as text, audios, animations and videos. Students were able to control the presentation by pressing stop, pause, continue, or volume on/off buttons. Figures 1 and 2 demonstrate two sample interfaces of the modules.

**Problem-solving tasks and interventions**

After studying each module, two problem-solving tasks were presented to the students. For each task, students had to choose the best solution among four possible choices. An example problem-solving task was stated as follows, “An object is moving at a constant speed. Its mass is 500 kg. Now an external force is pushing the object in a direction that is opposite to the object’s movement. As a result, the object stopped moving after five seconds. It is known that the amount of the external force is 2000 N. What was the speed of the object before it was pushed by the external force?” For the students in the PS/KR and PS/KCR groups, the system provided question prompts to scaffold their problem solving. The other two groups did not receive question prompts. Instead, they had to work on their own to solve each problem. The question prompts were designed by two science teachers and a researcher with expertise in problem solving. The questions were intended to guide students to strategically approach a problem instead of jumping quickly to solutions. The questions were based on the problem-solving processes identified in the literature, which include understanding the nature of a problem, planning a solution, and executing the plan (Jonassen, 2011; Polya, 1957). As an example, for the problem-solving task described above, three questions were asked: (a) What is the main purpose of this question? (c) In the problem, which of following do we already know? (b) Based on your answers to the two previous questions, which formula is applicable to the problem? Each of the questions had multiple answers for students to choose from.

Two types of feedback were provided to the participants as they worked on the problem-solving tasks. In the PS/KR and noPS/KR conditions, students received feedback about whether or not their answers were correct. Figure 3 illustrates an example feedback received by the PS/KR group. In the other two conditions (PS/KCR and noPS/KCR), students not only knew whether their answers were correct, but also knew the correct answer to each question. Figure 4 illustrates an example feedback received by the noPS/KCR group.
Knowledge acquisition test

The knowledge acquisition test was devised by two science instructors to ensure that the items adequately and appropriately assessed students’ mastery of the content in the instruction. The test was administered twice as pretest and posttest. The test contained 23 multiple-choice questions. Students received one point for each correct answer, and could earn up to 23 total points. Cronbach’s alphas for the pretest and posttest were .78 and .81, respectively, suggesting satisfactory reliability at each test time.

Cognitive load measure

The cognitive load measure was translated from Lin, Atkinson, Christopherson, Joseph, and Harrison (2013), which had been previously described and adopted by Gerjets, Scheiter, and Catrambone (2004). The instrument included three subjective questions to assess students’ cognitive load. Each question utilized a Likert-type rating scale from 1 (very low cognitive load) to 8 (very high cognitive load). The three questions were: (a) How difficult was the study? (b) How much mental effort did it take to learn from the materials? (c) How frustrated were you during the study? Students’ ratings of the three questions were averaged to represent their level of cognitive load, with 8 being the highest. Cronbach’s alpha for the measure was .85.
Procedure

In this study, after the students were assigned to the four treatment groups, each group participated in two 90-minute sessions per week for a period of two weeks. The study took place in computer labs with a teacher and three researchers present throughout the process. On the first day of the study, the participants were introduced to the research team, informed of the general purpose of the study, and then given a description of the procedure. After the orientation, students had approximately 20 minutes to individually complete the pretest.

Upon completion of the pretest, each student was given a unique username and password to access the e-learning environment and proceed with the learning modules. Depending on their treatment conditions, students had to solve problems with or without question prompts scaffolding, while receiving different types of feedback (KR or KCR). Upon their completion of all the five modules and problem-solving tasks, the students took the posttest and completed the cognitive load measure at the end of the second week.

Results

Table 1 shows the means and standard deviations of the knowledge acquisition pretest and posttest, as well as the cognitive load measure for the four treatment conditions. No statistical difference was found between the groups in the pretest, $F(3, 114) = 1.48, p > .05$, partial $\eta^2 = .04$. The posttest means ranged from 15.45 (noPS/KCR group) to 18.67 (PS/KCR group), suggesting that all the groups had some improvement from the pretest (ranged 10.00-11.00), but the scale was not very large. With regard to cognitive load, the means ranged from 3.30 (PS/KCR group) to 4.28 (noPS/KCR group), suggesting that the groups perceived low to medium levels of cognitive load.

<table>
<thead>
<tr>
<th></th>
<th>Prompts Provided (PS)</th>
<th>No Prompts (noPS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KR (N = 27)</td>
<td>KCR (N = 30)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Pretest</td>
<td>10.70 (2.20)</td>
<td>10.77 (1.68)</td>
</tr>
<tr>
<td>Posttest</td>
<td>16.15 (4.00)</td>
<td>18.67 (3.19)</td>
</tr>
<tr>
<td>CL</td>
<td>3.98 (1.52)</td>
<td>3.30 (1.50)</td>
</tr>
</tbody>
</table>

Effects on knowledge acquisition

A 2 x 2 ANCOVA was conducted to evaluate the effects of two question prompts conditions (PS and noPS) and two feedback types (KR and KCR) on students’ performance in the knowledge acquisition posttest, with the pretest scores as the covariate. The results of the ANCOVA indicated a significant main effect of question prompts, $F(1, 113) = 7.32, p < .01$, partial $\eta^2 = .06$, and a significant effect of question prompts by feedback interaction, $F(1, 113) = 5.56, p = .02$, partial $\eta^2 = .05$. The main effect for the feedback type was not significant. Given the significant interaction, interpretations of main effects were set aside in favor of interpretation of the simple main effects of question prompts and feedback respectively. The simple main effect of the question prompts was found to be significant within the KCR condition, $F(1, 113) = 13.38, p < .001$, partial $\eta^2 = .11$, with the PS group outperforming the noPS group. Within the KR condition, there was no significant difference between the PS and noPS conditions. With regard to the simple main effect of feedback, there was a significant difference within the PS condition, $F(1, 113) = 7.06, p < .01$, partial $\eta^2 = .06$, with the KCR group outperforming the KR group. No significant difference was found between KR and KCR within the noPS condition. Figure 5 illustrates the four group’s knowledge acquisition performance.
Effects on cognitive load

A 2 x 2 ANOVA was conducted to evaluate the effects of question prompts and feedback on students’ perceived cognitive load. The results indicated a significant prompts by feedback interaction, $F(1, 114) = 3.93, p = .0498$, partial $\eta^2 = .033$. The main effects for both feedback and question prompts were not significant. This interaction effect was further investigated by conducting simple main effects of question prompts within each feedback condition. There was a significant difference within the KCR condition, $F(1, 114) = 6.72, p = .011$, partial $\eta^2 = .11$, with the PS group reporting lower cognitive load than the noPS group. The cognitive load of the four groups is illustrated in Figure 6.

Figure 5. Estimated marginal means of knowledge acquisition posttest by groups

Figure 6. Estimated marginal means of cognitive load by groups
Discussion

This study investigated the effects of problem-solving question prompts and feedback on students’ knowledge acquisition and perceived cognitive load as they studied a web-based module on Newtonian mechanics. A particular purpose of the study was to bring together two lines of research, scaffolding research and cognitive load theories. While both research areas are related to web-based multimedia learning, few studies have empirically examined the effects of scaffolding on cognitive load.

Question prompts, feedback, and knowledge acquisition

Aligned with the existing literature (e.g., Bulu & Pedersen, 2010; Ge et al., 2005; Kim & Hannafin, 2010; Raes, Schellens, Wever, & Vanderhoven, 2012), this study found that the problem-solving question prompts led to improved knowledge acquisition. Moreover, the effect of question prompts was more pronounced when combined with the KCR feedback. While past studies on scaffolding and question prompts examined their use to support ill-structured problem solving, this study implemented question prompts to support students’ solution of conceptual, application problems that usually have convergent answers and preferred solution paths (Jonassen, 2000; 2011). When solving this type of problems, a human tutor in face-to-face settings can direct students’ attention and guide them through the process. In web-based learning, however, such scaffolding is often not available. Students usually receive feedback after they solve a problem. The findings from this study suggest that question prompts can help students better understand a problem and plan for solutions to the problem, which can lead to improved knowledge acquisition. The strengths of question prompts can be further enhanced when combined with KCR feedback.

With regard to the use of the two types of feedback, KR and KCR, the findings are different from those in the past studies (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Jaehning & Miller, 2007; van der Kleij et al., 2012). In the existing literature, feedback is often provided after students complete certain problem-solving tasks. This study took a different approach by integrating feedback with problem-solving question prompts prior to students’ solution of problems. In such a context, this study found that the types of feedback did not make a significant difference on students’ knowledge acquisition overall. However, the feedback did make a difference when problem-solving question prompts were available. That is, KCR can lead to better knowledge acquisition than KR when combined with problem-solving question prompts.

Question prompts, feedback, and cognitive load

In this study, neither question prompts nor the types of feedback had a general impact on students’ cognitive load. With regard to the cognitive load of KR and KCR, while the comparison of the two feedback types was not previously studied in the problem-solving context, the findings indicate that the two types of feedback generally did not cause difference in cognitive load. Regarding the problem-solving question prompts’ lack of impact on cognitive load, one possible explanation could be that, while the prompts might help to reduce extraneous load, students’ germane cognitive load might have been increased as they attempted to answer the questions. The overall lack of difference in cognitive load among the four groups might be partially attributed to the students’ relatively high level of prior knowledge which is shown in the knowledge acquisition pretest. With certain levels of prior knowledge, the students might not feel the study as taxing as might those with little prior knowledge, which could have led to the relatively low cognitive load and lack of difference among the treatment groups.

On the other hand, for the two groups that received corrective feedback, problem-solving question prompts did help to reduce cognitive load. It could be reasoned that for the students who did not receive question prompts, when they learned the correct answer to a problem following an incorrect response, they might have wondered why, even though they were not prompted to do so. Without previous question prompt scaffolding, such reasoning process might have taken this group of students extra mental effort and even caused frustration. The finding suggests that when learners take time to go over necessary problem-solving mental processes by responding to question prompts and receiving corrective feedback, they may be able to manage their cognitive resources more effectively, which also contributes to improved knowledge acquisition.
This study contributes to the cognitive load and scaffolding literature by providing empirical evidence against the intuitive assumption that scaffolding could reduce cognitive load during problem solving. It is possible that scaffolding, such as question prompts, does not help to reduce cognitive load in problem solving. While scaffolding in itself may not have an effect on cognitive load, it may have an effect when combined with other instructional interventions such as corrective feedback. Therefore, in examining the relationship between scaffolding and cognitive load, it is productive to go beyond a one-dimensional, universal-effect model to investigate how scaffolding influences cognitive load in different instructional contexts.

Conclusion and implications

Overall, the findings suggested that question prompts, when combined with corrective feedback, can provide an effective means of supporting problem solving and knowledge acquisition in web-based science learning. Moreover, such a combination can not only lead to improved knowledge acquisition, but also to reduced cognitive load which may further contribute to improved learning and reduced stress. The study offers several practical implications. First, when students’ learn to apply principles and rules to solve science problems in web-based learning, feedback alone is not sufficient. Scaffolding such as question prompts is necessary to engage them in the thinking processes that lead to problem solution. Second, in the web-based learning context, use the combination of question prompts and KCR to provide optimal support for students’ problem solving and maximize students’ use of cognitive resources on tasks directly related to learning.

This study has a few limitations. First, the same knowledge acquisition test was administered twice as pretest and posttest. Although the two tests were two weeks apart, during which students studied five learning modules, the pretest could have still had a learning effect on the students in their study. Second, although the problem-solving question prompts and the actual problem-solving tasks were different content-wise, they both took a multiple-choice format. Therefore, the PS groups could have potentially gained more familiarity with multiple-choice questions than the noPS groups, which could have partially led to the PS groups’ better performance and lower perceived cognitive load.

Future studies could use parallel forms for pretest and posttest, and adopt additional question format in measuring students’ knowledge acquisition. In addition, future studies should examine the effects of question prompts and feedback on specific types of cognitive load (DeLeeuw & Mayer, 2008). Further, future studies can also incorporate elaborated feedback, another frequently used feedback type, to examine its impact on learning and cognitive load in web-based science problem solving. Finally, it is worthwhile to study how scaffolding interacts with other instructional interventions to impact cognitive load.

References


Metacognitive Load – Useful, or Extraneous Concept? Metacognitive and Self-Regulatory Demands in Computer-Based Learning

Rolf Schwonke
Department of Psychology, University of Freiburg (Germany), Freiburg, Germany // schwonke@psychologie.uni-freiburg.de

ABSTRACT
Instructional design theories such as the cognitive load theory (CLT) or the cognitive theory of multimedia learning (CTML) explain learning difficulties in (computer-based) learning usually as a result of design deficiencies that hinder effective schema construction. However, learners often struggle even in well-designed learning environments. In this theoretical paper, I will argue that cognitive resources-oriented theories such as the cognitive load theory might profit from extending their predominantly cognitive focus to one that additionally considers metacognitive and self-regulation demands. Empirical results on learning from multiple external representations and research on tool use are integrated to illustrate that computer-based learning environments usually pose a variety of cognitive, metacognitive and self-regulatory demands on learners which require knowledge and skills that learners often lack. Specifically, empirical findings suggest that most learners are unable to regulate their learning automatically. I thus argue that these activities consume working-memory resources as do activities that are closely related to schema construction. My article concludes with suggestions on how the concept of metacognitive load can be incorporated into resource-oriented theories such as the CLT to explain a wider variety of phenomena.

Keywords
Cognitive load, Computer-based learning, Metacognition, Self-regulated learning, Working memory

Introduction
The environment(s) we live in are becoming ever more complex (Field, 2006). To meet the demands of contemporary and future learning environments and work places, the complexity of curricula, learning environments, and learning tasks has increased accordingly (van Merriënboer & Swijinmans, 2009). In this context, complexity is not something we should try to avoid in general. Certain levels of complexity are actually necessary to stimulate and foster higher-order cognitive and metacognitive processes (Sawyer, 2006). However, there is a challenge as well: Increasingly complex and rapidly changing learning demands encounter the relatively stable and limited cognitive equipment of humans. Increasingly powerful computers and high-resolution computer screens make it possible to display ever more complex arrangements of texts, pictures, animations, and sounds simultaneously.

In an ideal world, learners would make flexible use of such complex arrangements and blend new information together with their background knowledge into increasingly rich mental models. Facing difficulties or impasses, learners would actively pose relevant questions, persistently hunt for answers, critically evaluate the quality of the answers retrieved, construct deep explanations of the subjective matter, apply the explanatory content to difficult problems, and consciously reflect on these cognitive activities (Graesser, McNamara, & VanLehn, 2005, p. 231). Empirical results, however, reveal several problems that learners usually encounter in rich learning environments. For example, they often have difficulty extracting and integrating information from complex displays (e.g., Ainsworth, 2006). Learners often have problems interacting effectively with the multifarious options that rich media environments usually offer (e.g., Clarebout & Elen, 2007). Related to the interaction problems, learners frequently have trouble effectively monitoring and regulating their learning activities (e.g., Azvedo, 2002). Such findings suggest that we need to balance the degree of complexity that is technically feasible with what is cognitively achievable and, therefore, educationally desirable.

A general question that arises is whether and how humans’ capabilities can keep pace with these rapidly changing demands. In the context of computer-based learning environments (CBLEs), a related, but more specific set of questions arises, namely how effectively can learners utilize the various features that CBLEs offer nowadays to acquire knowledge and skills in different domains (e.g., mathematics or biology)? What difficulties do learners typically encounter when interacting with CBLEs? How are such difficulties explained by contemporary psychological theories? What cannot be explained well and where should we, therefore, modify, extend or combine contemporary theories? And finally, how can CBL be made even more effective?
The overall goal of this article is to contribute to our understanding of factors that can account for the typical difficulties learners encounter in CBL. In the following I argue that, especially in individual learning in front of a computer screen (which is becoming and will remain an increasingly more common and important learning setting nowadays), successful learning depends primarily on learning environments that support not only learners’ cognitive, but their metacognitive and self-regulation activities as well. Although there is growing interest in supporting all kinds of such activities, three factors are often overlooked. First, unless they are fully automatized, all these activities compete for limited mental resources. Second, as with cognitive demands that do not always benefit learning, some metacognitive and self-regulatory demands might not always be beneficial either. Third, learners will or cannot always take full advantage of the support offered. The latter will usually initially depend on different types and degrees of expertise on behalf of the learners, i.e., their domain-specific prior knowledge and cognitive, metacognitive, and self-regulatory skills. My central argument is that with their growing complexity, CBLs pose cognitive as well as metacognitive and self-regulation demands on learners. I further argue that these demands may sometimes be beneficial, sometimes irrelevant (but neutral), but occasionally also harmful for learning. To support these arguments, I examine typical demands that CBLs pose on learners and discuss the role of their expertise in coping with these demands. Against the background of empirical results from research on learning from multiple external representations (of the learning contents) and research on tool use (i.e., learners’ use and conceptions of support measures in CBLs) I finally propose incorporating the notion of metacognitive demands in resource-oriented theories such as such as the cognitive load theory (CLT; Chandler & Sweller, 1991; Sweller, 2003; Sweller, 2004) and the cognitive theory of multimedia learning (CTML; Mayer, 2005).

The remainder of the text is structured as follows. First, I sketch how CLT and CTML contribute to the explanation of many (but not all) difficulties in CBL. I then illustrate how external representations and support measures (often referred to as “tools”) as the constituent elements of CBLs pose different demands on learners and how learners’ levels of expertise are related to this. Against this background, I argue that metacognitive theories, and especially theories of self-regulated learning (e.g., Winne, 1996) can provide valuable complementary explanations of difficulties in CBL. Finally, I demonstrate how resource-oriented theories such as CLT might profit from incorporating the notion of metacognitive load.

A cognitive-resources perspective on learners’ difficulties in CBL

Instructional-design theories such as the cognitive load theory and/or the cognitive theory of multimedia learning explain suboptimal learning outcomes mainly as a consequence of less-than-ideal learning-environment designs; design deficiencies that absorb too much of learners’ cognitive capacities (or resources), leaving too little capacity for making sense of the material and for learning activities. From the perspective of CLT and CTML, learning environments should therefore be constructed in ways compatible with the human cognitive architecture. According to these approaches, the limitations of human working memory in particular (Baddeley, 1992) should be considered when designing learning environments and learning tasks. As such, CLT and CTML can be called cognitive-resource oriented theories. In these resource-oriented theories (as I refer to them in the remainder of this article), working memory is the central hypothetical structure of the human cognitive system where incoming information from the sensory channels (mainly visual and auditory; Mayer 2005) is temporarily stored and processed.

In CLT, the critical learning mechanisms are schema acquisition and schema automation. It is assumed that novel information is processed in the light of existing schemata that allow the classification of objects and problems into categories and that provide the necessary information on how to deal with novel information (Sweller, 1994). In CTML, critical learning mechanisms refer to the selection of relevant (verbal and visual) incoming information, the organization of that novel information into meaningful structures and, most importantly, the integration of novel with existing information into coherent mental representations to be stored in long-term memory (LTM). CTML also assumes that information is stored in the form of schemata. Integration, therefore, refers to the construction of novel schemata and the modification of existing schemata. Integration can thus be equated with learning, given that the integration results in a lasting change in LTM. In CLT, an initial form of working-memory load can be attributed to the learning content’s difficulty. Basically, this intrinsic load is related to the number of relevant concepts in a learning content that must be maintained and processed simultaneously in working memory (i.e., element interactivity) in order to understand the content. Intrinsic load varies with the learners’ prior knowledge and experience in given and related domains (e.g., Kalyuga, 2011). Prior knowledge – in form of highly integrated and hierarchically structured schemata – allows learners to treat multiple related concepts as single chunks of information.

The overall goal of this article is to contribute to our understanding of factors that can account for the typical difficulties learners encounter in CBL. In the following I argue that, especially in individual learning in front of a computer screen (which is becoming and will remain an increasingly more common and important learning setting nowadays), successful learning depends primarily on learning environments that support not only learners’ cognitive, but their metacognitive and self-regulation activities as well. Although there is growing interest in supporting all kinds of such activities, three factors are often overlooked. First, unless they are fully automatized, all these activities compete for limited mental resources. Second, as with cognitive demands that do not always benefit learning, some metacognitive and self-regulatory demands might not always be beneficial either. Third, learners will or cannot always take full advantage of the support offered. The latter will usually initially depend on different types and degrees of expertise on behalf of the learners, i.e., their domain-specific prior knowledge and cognitive, metacognitive, and self-regulatory skills. My central argument is that with their growing complexity, CBLs pose cognitive as well as metacognitive and self-regulation demands on learners. I further argue that these demands may sometimes be beneficial, sometimes irrelevant (but neutral), but occasionally also harmful for learning. To support these arguments, I examine typical demands that CBLs pose on learners and discuss the role of their expertise in coping with these demands. Against the background of empirical results from research on learning from multiple external representations (of the learning contents) and research on tool use (i.e., learners’ use and conceptions of support measures in CBLs) I finally propose incorporating the notion of metacognitive demands in resource-oriented theories such as such as the cognitive load theory (CLT; Chandler & Sweller, 1991; Sweller, 2003; Sweller, 2004) and the cognitive theory of multimedia learning (CTML; Mayer, 2005).

The remainder of the text is structured as follows. First, I sketch how CLT and CTML contribute to the explanation of many (but not all) difficulties in CBL. I then illustrate how external representations and support measures (often referred to as “tools”) as the constituent elements of CBLs pose different demands on learners and how learners’ levels of expertise are related to this. Against this background, I argue that metacognitive theories, and especially theories of self-regulated learning (e.g., Winne, 1996) can provide valuable complementary explanations of difficulties in CBL. Finally, I demonstrate how resource-oriented theories such as CLT might profit from incorporating the notion of metacognitive load.

A cognitive-resources perspective on learners’ difficulties in CBL

Instructional-design theories such as the cognitive load theory and/or the cognitive theory of multimedia learning explain suboptimal learning outcomes mainly as a consequence of less-than-ideal learning-environment designs; design deficiencies that absorb too much of learners’ cognitive capacities (or resources), leaving too little capacity for making sense of the material and for learning activities. From the perspective of CLT and CTML, learning environments should therefore be constructed in ways compatible with the human cognitive architecture. According to these approaches, the limitations of human working memory in particular (Baddeley, 1992) should be considered when designing learning environments and learning tasks. As such, CLT and CTML can be called cognitive-resource oriented theories. In these resource-oriented theories (as I refer to them in the remainder of this article), working memory is the central hypothetical structure of the human cognitive system where incoming information from the sensory channels (mainly visual and auditory; Mayer 2005) is temporarily stored and processed.

In CLT, the critical learning mechanisms are schema acquisition and schema automation. It is assumed that novel information is processed in the light of existing schemata that allow the classification of objects and problems into categories and that provide the necessary information on how to deal with novel information (Sweller, 1994). In CTML, critical learning mechanisms refer to the selection of relevant (verbal and visual) incoming information, the organization of that novel information into meaningful structures and, most importantly, the integration of novel with existing information into coherent mental representations to be stored in long-term memory (LTM). CTML also assumes that information is stored in the form of schemata. Integration, therefore, refers to the construction of novel schemata and the modification of existing schemata. Integration can thus be equated with learning, given that the integration results in a lasting change in LTM. In CLT, an initial form of working-memory load can be attributed to the learning content’s difficulty. Basically, this intrinsic load is related to the number of relevant concepts in a learning content that must be maintained and processed simultaneously in working memory (i.e., element interactivity) in order to understand the content. Intrinsic load varies with the learners’ prior knowledge and experience in given and related domains (e.g., Kalyuga, 2011). Prior knowledge – in form of highly integrated and hierarchically structured schemata – allows learners to treat multiple related concepts as single chunks of information,
thereby reducing processing demands and facilitating further schema acquisition. With further exposition to similar situations or problems where acquired schemata have been successfully applied (i.e., by practice), learners will ultimately be able to solve such problems effortlessly because the required knowledge and routines have been automated. In CLT and CTML it is further assumed that schema acquisition can only succeed when the working memory is not already overburdened by processing demands that do not contribute to learning. CLT refers to such unnecessary processing demands on the learning environment as extraneous load (Chandler & Sweller, 1991). CTML refers to such demands as incidental processing (Mayer 2005). In order to construct, refine, and automate schemata CLT and CTML assume that learners must engage in conscious cognitive processes (Paas et al., 2003). In the original CLT framework, such processes are assumed to induce a third form of (desirable) working-memory load referred to as germane cognitive load (Sweller, van Merriënboer & Paas, 1998). CTML refers to such demands that support schema-construction (i.e., selection, organization, and integration) as essential processing. However, even when a CBLE is designed to induce an appropriate amount of intrinsic load (relative to learners’ prerequisites), a minor amount of extraneous load and fair amount of germane load, the environment’s effectiveness will also strongly depend on the willingness of learners to engage in relevant cognitive processes. In acknowledging this active role of learners, the concept of mental effort was introduced. Mental effort is defined as the amount of cognitive capacity that a learner allocates to accommodate the demands imposed by the task (Paas et al., 2003).

Resource-oriented frameworks have greatly expanded our understanding of relevant factors that influence learning. For example, CLT research has identified crucial task characteristics such as task format, task complexity, time pressure, or the use of multimedia (Paas et al., 2003). However, these task characteristics also reveal the two potential shortcomings of resource-oriented theories elaborated below.

**Potential shortcomings of resource-oriented theories**

An initial potential shortcoming involves the coarse or broad terms of resource-oriented theories. CLT and CTML are framework theories. As such, these theories operate on a relatively high level of abstraction. Due to this high level of abstraction, these theories’ descriptions are necessarily rather broad, often leaving ample room for (post-hoc) interpretation. On the other hand, these theories are also instructional design theories. As such, they aim to inform instructional designers, teachers, and educational decision makers. To be (even) more informative, such theories should probably provide explanations on a more specific level. For example, in CLT the types of activities that may induce germane load are not elaborated further, making it difficult to design activities that foster germane from a cognitive load perspective. In an evaluation of CLT, Schnotz and Kürschner (2007) define germane load as a kind of working memory load “due to cognitive activities that aim at intentional learning and that go beyond simple task performance” (p. 496). Among the cognitive activities that Schnotz and Kürschner (2007) assume to induce “germane load” are, for example, the conscious search for patterns in the learning material (i.e., mindful abstraction), the conscious application of learning strategies, the restructuring of a problem representation (in order to solve a task more easily), and metacognitive processes that monitor cognition and learning.

From my perspective, two things are important to keep in mind. First, there might be more activities or processes relevant to effective learning beyond those proposed by resource-oriented theories. Second, several such activities seem to require conscious effort (at least until fully automated) and can thus induce working memory load. Initial candidates for such learning-relevant activities are metacognitive and self-regulatory activities. However, instead of assuming that such activities would per se induce germane load, I prefer to argue that such activities can induce either germane load or extraneous load (or none at all) depending primarily on the learner’s experience with that activity. For example, this assumption is supported by the well-known fact that the acquisition of a learning strategy initially usually decreases learning performance and outcomes (referred to as the “valley of tears”) until that strategy can be applied automatically (i.e., without much conscious effort). According to this view, metacognitive and self-regulatory activities can be beneficial, ineffective, or even detrimental for schema construction. In this context we can ask whether there are any other unnecessary or even harmful demands on working memory that may have been overlooked or ignored by resource-oriented theories and if so, what the sources of these additional demands are.

In order to answer these questions, I propose to seek variables that resource-oriented theories do not usually focus on. First of all, resource-oriented theories are much more deeply concerned with cognitive demands that directly affect schema construction than with metacognitive and self-regulative demands and the potential roles that these processes might play in learning. Furthermore, resource-oriented theories assume well-integrated (domain-specific) schemata
in long-term memory to guide metacognitive processes such as planning, monitoring, or evaluating rather automatically. Being automatic, these processes would not consume any working-memory resources (e.g., Sweller et al., 2011). However, how can we be certain that these processes consume no capacity? What, for example, about false beliefs or misconceptions about the purpose(s) of a CBLE or their various features? When such misconceptions (stored as schemata in long-term memory) guide the learning process, it is reasonable to assume that such misleading schemata would impose extraneous demands on learners’ cognitive systems.

In addition, resource-oriented theories potentially (lead researchers and practitioners to) underestimate the complexity of the relationships between task characteristics, patterns of cognitive load, and learning outcomes. Can we expect high-quality learning outcomes to result from a specific “state-of-the-art” design rather deterministically? From the CLT perspective, the answer would probably be “yes.” Given a specific instructional design, CLT would predict specific types of activities by the learners which should consequently result in specific patterns of cognitive load (Gerjets & Scheiter, 2003). However, Gerjets and Scheiter, for example, demonstrated that variables such as the goal configurations of teachers and learners, and learners’ processing strategies can moderate the effects of instructional design on extraneous load. In their study, learners’ beliefs and expectations influenced goals and intentions. Goals and intentions then strongly determined strategy choice and the mental effort invested (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). The resulting activities are assumed to correspond relatively straightforwardly to specific patterns of cognitive load, which then predict learning outcomes (Gerjets & Hesse, 2004). Such results suggest a more complex pattern of relationships among task, load, and outcomes. From a resource-oriented perspective, important learner activities or processes might also be overlooked as the following example will show.

In an evaluation study of a tried-and-tested PACT Geometry Cognitive Tutor (Aleven & Koedinger, 2000), students displayed significant improvement from pretest to posttest (leaving room for further improvement). The Cognitive Tutor program provided tailored, just-in-time support to the students during problem-solving. The students had ample experience with the CBLE and the learning domain. From a cognitive-load perspective, such a result would probably be interpreted as being the result of a successful combination of just enough intrinsic load (relative to students’ level of prior knowledge), a moderate amount of extraneous amount (due to just-in-time support offered during problem-solving), and a reasonable amount of germane load (due to task design, the support offered, and the mental effort the learners invest). However, an in-depth log-file analysis revealed inefficient learning behavior. Learners waited too long to ask for help in case of errors, and they largely ignored a glossary that provided definitions and examples of the geometry principles. Moreover, when learners did ask for help, they immediately looked for the most specific level of help that virtually provided the solution to the task at hand. The authors attributed this inefficient help-seeking behavior to a lack of metacognitive knowledge and skills (e.g., being able to monitor one’s progress) and proposed training these skills so that learners can make more effective use of available help. It’s probably fair to say that in this particular study, neither cognitive load nor mental effort was assessed. However, even when learners might have reported little invested mental effort, without detailed information on their actual learning and help-seeking behavior, being aware that learners had not invested enough effort would probably not provide us with enough information for any instructional intervention.

As detailed later, theories of metacognition and self-regulation can provide us with alternative and complementary notions and models. In the next sections, empirical results of research on learning from multiple external representations and tools are reported to show that CBLEs can pose quite heavy metacognitive and self-regulation demands on learners in numerous ways.

**Sources of metacognitive and self-regulatory demands in computer-based learning**

To understand why resource-oriented theories cannot fully account for learners’ difficulties in CBL, it is worth looking at the constituent elements of contemporary computer-based learning environments. A CBLE has to include different types of information resources (e.g., texts, illustrations, and help facilities). These resources can be functionally differentiated into resources representing the subject matter (e.g., principles of geometry) and resources supporting the subject matter’s acquisition (i.e., help facilities or so-called tools such as a glossary). Both types of resources are often constructed of more than one (often multiple) external representations related to the same concept. For example, a geometry word problem might be accompanied by a diagram showing known and unknown angles as
described in the word problem. Likewise, the definition of a geometry principle in a glossary might be illustrated by a corresponding diagram.

As such, multiple external representations and embedded tools can be regarded as the generic building blocks of computer-based learning environments, but also as two main sources of (potentially non-beneficial) metacognitive and self-regulatory demands. The demands increase with the number and qualities of relationships between these variously-constituent types of elements to each other as well as to the learning task. Closely related to this increase in complexity (comparable to the element interactivity within the learning content) are the demands on learners’ self-regulation. For example, the tactical decisions concerning the allocation of attentional and cognitive processes become more demanding with each additional element (Lajoie, 1993).

Research on learning from multiple external representations (e.g., Ainsworth, 2006; van Someren, Reimann, Boshuizen, & de Jong, 1998) and research on tool use (e.g., Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Clarebout & Elen, 2008) provide us with many empirical examples of the typical difficulties learners experience in complex environments, and of more and less successful “coping” strategies. In the next two sections I illustrate results from these two lines of research which reveal that the difficulties in CBL can be differentiated as cognitive, metacognitive and self-regulatory difficulties.

**Metacognitive and self-regulatory demands of external representations**

Multiple external representations are thought to fulfill (at least) three different (but related) cognitive functions (Ainsworth, 2006). First, multiple representations can complement each other and thus provide a more complete picture of a difficult concept. For example, a verbal description of a mathematical function (e.g., \( y = x^2 - 2 \)) can be accompanied by a line drawing depicting that function. Second, multiple representations can help to constrain each other’s interpretations. For example, a table with data can be accompanied by a scatterplot. Inspecting the form of the points as detailed in the scatterplot can constrain how learners interpret the data from the table. Finally, and probably most importantly, multiple representations can be integrated (by the learners) to construct a more abstract internal representation of the material presented externally. For example, from a scatterplot and a table presented together, learners can infer a general rule about a functional relationship between the depicted data.

Research on learning from multiple external representations focuses mainly on how learners make sense of different symbol systems (e.g., text, numbers, and realistic pictures) and how external representations can be combined to contribute most to understanding and learning. Empirical results show that learners often have trouble learning from multiple external representations (e.g., Ainsworth, Bibby, & Wood, 2002). Learners tend to use different external representations in isolation and to use only a sub-sample of available representations, even when a learning task strongly suggests attending to all the representations available. Learners seem to find constructing referential connections between the concepts depicted by different external representations particularly difficult (Ainsworth, 2006).

In summary, students have difficulty (a) relating the contents of different external representations to one another and (b) understanding how different external representations can contribute to understanding and learning (i.e., understanding the representations’ didactic functions). Difficulty relating contents of different representations reflects either cognitive or self-regulation problems. Having trouble understanding the didactic function of different (types of) external representations probably reflects metacognitive knowledge deficits more than cognitive deficits.

**Metacognitive and self-regulatory demands of support measures (tools)**

CBLEs can have a variety of overarching purposes, for example helping learners gain deeper understanding, acquire knowledge or develop skills quicker, or learn at one’s own pace. As these overarching purposes require a number of more specific didactical strategies or tactics, a number of smaller, more specific help facilities are usually devised that are integrated in a package that we call a CBLE. From a functional perspective, therefore, we can consider CBLEs as collections of more or less related (or integrated) support measures (i.e., tools), each measure serving a specific purpose or function. It is thus worth taking a look at the functions that different tools should fulfill in the
eyes of the instructional designers, how learners conceive the purpose of different tools, how learners actually use tools, and, finally, how the use of such tools might affect learning.

Tools in CBLEs are artifacts designed to support cognitive and metacognitive processes related to the actual learning task. An example of a cognitive tool is a pocket calculator embedded in a CBLE for algebra or geometry. An example of a metacognitive tool would be some sort of overview (e.g., a table) of accomplished and open tasks (to facilitate monitoring and planning). Tool use can thus be defined as student-system interactions (with help facilities in CBLEs) that aim to overcome or prevent problems during learning (Aleven & Koedinger, 2000).

Research on tool use has described typical ways in which learners use available tools and how different uses affect learning outcomes. A typical finding is that learners often ignore available tools even after they have proven to be useful (Clarebout & Elen, 2007). In addition, learners often use tools inadequately or at least not as intended by the instructional designers (e.g., Aleven et al., 2003; Clarebout & Elen, 2006). For example, in a log-file analysis on how school children in a geometry course used different tools of an intelligent tutoring system (a Cognitive Tutor Geometry), students did not use errors as a signal to request for help (Aleven and Koedinger, 2002). Moreover, when students did decide to request help, they often proceeded straight to the most solution-specific hints. This so-called bottom-out hint strategy indicates that at least some learners tend to use available help facilities in a non-learning-oriented manner (Aleven & Koedinger, 2002). Such gaming-the-system behavior is usually negatively associated with learning outcomes (Baker, Corbett, & Koedinger, 2004). On the other hand, for those learners who took their time to study the bottom-out hints, learning outcomes increased (Shih, Koedinger, & Scheines, 2008).

To summarize: learners tend to avoid using tools, they do not use them early enough or not at the right occasion. In addition, some learners tend to use tools too often or for the “wrong” purpose. Obviously, learners have difficulty while learning to map impasses to available help facilities. In other words, they have trouble deciding when to refer to which type of support. All these examples of typical ways of using tools in CBLE reflect metacognitive or self-regulation deficits more than cognitive deficits.

This overall negative picture might suggest that providing metacognitive and self-regulation support in CBLE is perhaps not very effective. In that case it would not be advisable for students to use such offers. There is, however, ample evidence that such support can be effective. For example, the frequency of monitoring activities has often been found to be a predictor of student learning (e.g., Winne, 2001; Winne & Hadwin, 1998). This could also be shown for studying complex topics using a hypermedia environment (Azevedo, 2005). For example, (Greene & Azevedo, 2007) could show that monitoring processes had a significant relation with the odds of having a more sophisticated mental model of a complex biological system (i.e., blood, heart, and circulatory system) at a knowledge posttest. Noticeably, the effect of monitoring processes on performance was above and beyond the effects of prior knowledge and age. Self-explanation is often considered as a metacognitive learning strategy (e.g., Aleven & Koedinger, 2002; Conati & VanLehn, 2000; see also Renkl, Berthold, Große, & Schwanke, 2013). Berthold and Renkl (2009) prompted self-explanations in learning probability theory in a CBLE (N = 170 high-school students; mean age approx. 16 years). They found that prompting self-explanation fostered conceptual understanding of the probability principles. Low-prior knowledge students, however, profited less from the intervention. The authors concluded that there are boundary conditions to be considered. Self-explanation prompts can lead to negative effects if the learners are confronted with learning materials that are very complex in relation to their prior knowledge (Berthold and Renkl, 2009). In a log-file analysis of a Cognitive Tutor study on learning geometry (Otieno, Schwanke, Renkl, Azevedo, & Salden, 2011) learning outcomes had stronger relationships to self-explanation performance than to problem-solving performance during learning. Additionally, self-explanation performance was a stronger predictor for learning outcomes than prior knowledge. These results suggest that metacognitive support can enhance learning provided that some boundary conditions are considered.

Overall, research on learning from multiple external representations and tool use highlight the many difficulties that learners have using these external resources effectively. Most findings suggest that learners’ problems are more closely related to regulation of the learning process than to schema construction itself. We can therefore speak of the metacognitive and self-regulation demands that learning environments pose on learners. To localize or isolate factors that might affect these demands, we need to take a closer look at metacognitive knowledge prerequisites and their self-regulatory skills. However, as we know that these skills are closely related to learners’ domain-specific knowledge (i.e., their level of expertise in a domain) the inter-relationships between metacognition, self-regulated learning and level of expertise are the topic of the next section.
The role of learners’ levels of expertise

Domain-specific knowledge (i.e., prior knowledge) is the best predictor of further learning (Ausubel, Novak, & Hanesian, 1978; Novak, 1990). Individuals’ prior knowledge and learning trajectories (i.e., the speed of learning), however, vary substantially. Such inter-individual differences in prior knowledge and trajectories can alter the effectiveness of instructional measures (known as the Expertise-Reversal-Effect; Kalyuga, Ayres, Chandler, & Sweller, 2003) as well as the effectiveness of individual learning activities. For example, prior knowledge can affect (whether and) how effectively learners process external representations (e.g., Wood & Wood, 1999). It can also affect learners’ need for help and how strategically they ask for help (e.g., Renkl, 2002). In a study on tool use in the domain of middle-school mathematics (Wood & Wood, 1999), low-prior knowledge learners used an intelligent tutoring system’s tools more frequently than high-prior knowledge learners. Seeking help after an error was associated positively with learning outcomes in low-prior knowledge learners but not in those with high prior knowledge. On the other hand, errors during learning were negatively related to learning outcomes only in low-prior knowledge learners. The authors concluded that encouraging low-prior knowledge learners - especially those who refuse to ask for help spontaneously - to seek help more strategically (i.e., not necessarily more often) might be a promising approach to reduce confusion and enhance learning performance. Similarly, Baker et al. (2004) found that especially low-prior knowledge learners used tools in a non-learning oriented way, which was negatively related to learning outcomes.

Prior knowledge provides the context for the interpretation of new information and, as such, also provides the background for any metacognitive considerations on the learners’ side (e.g., deciding on the need for help). Generally, metacognitive knowledge and regulation have been found to improve together with expertise within a particular domain. However, although domain-specific knowledge can facilitate the acquisition and use of metacognition, high levels of domain knowledge do not guarantee high levels of metacognition (Schraw, 1998).

Low prior knowledge can also impede metacognitive functioning (e.g., judging whether another hint would help to overcome an impasse; Schraw, 1998). Low-prior knowledge students have, for example, been found to use available information resources sub-optimally, which was attributed to metacognitive deficits (e.g., Baker et al., 2004). Low-prior knowledge learners are also more generally dependent on structure and scaffolding than are high-prior knowledge students (e.g., Kalyuga, 2007; Renkl, Stark, Gruber & Mandl, 1998). Therefore, low-prior knowledge learners might need metacognitive support more than high-prior knowledge learners. Often, however, low-prior knowledge learners do not benefit from metacognitive support. This finding can be explained by the fact that understanding and using a metacognitive strategy (in real time) is a cognitively-demanding activity. Therefore, especially low-prior knowledge learners can easily be overwhelmed, for example when advised to follow prescribed rules on when to use certain help facilities. The instructional challenge is to implement metacognitive support in a cognitively manageable way.

In summary, prior knowledge affects how much learners depend on information resources and how effectively they can make use of them. Prior knowledge can also affect how much learners profit from cognitive and metacognitive support. Against this background I hope to illustrate how metacognitive theories and self-regulation theories can contribute to our understanding of difficulties in CBL, and how these theories and resource-oriented theories might be integrated to provide a more complete picture of potentials and difficulties in CBL.

A self-regulated-learning perspective on difficulties in CBL

Self-regulated learning can be described as a behaviorally, metacognitively, and motivationally active participation in one’s own learning (Zimmerman, 1986). Self-regulated learners employ cognitive strategies (e.g., elaboration) to attain learning goals. Choice of strategies, their application, and the quality of the outcomes of strategy application are embedded within and controlled by metacognitive activities such as planning, monitoring, and self-evaluating (Zimmerman, 1990).

Moreover, as with many CBLEs, learners can largely decide on their own whether, when and how to make use of available resources, one can argue that a self-regulated perspective (e.g., Schiefele & Pekrun, 1996; Boekaerts, 1999; Winne, 1996; Winne & Perry, 2000) becomes increasingly important in CBL. Even in highly structured environments (e.g., those that structure the sequence of learning tasks), room remains for variation in learners’
choices and execution of cognitive and metacognitive activities. To make effective use of available external resources, learners need to adequately allocate and regulate their attentional and cognitive resources during learning. However, with each additional external resource, tactical decisions as to where and when to use one or the other resource become more demanding (Lajoie, 1993). Multiple sources of information can, therefore, easily overwhelm learners’ self-regulatory capacities (Ainsworth et al., 2002).

From a metacognitive and self-regulatory perspective, learners should therefore be equipped with internal resources that allow them to cope with such additional metacognitive and self-regulatory demands. One such important internal resource is learners’ knowledge about available external resources. Such knowledge can be conceptualized as a subtype of metacognitive knowledge. As the knowledge about factors that affect cognitive activities, meta-cognitive knowledge (Flavell, 1979) refers to three broad categories: (a) person, (b) task, and (c) strategies. In Flavell’s classic definition, the “task” category includes information about a proposed task available to someone, including knowledge about tangible resources necessary for task completion. As such, knowledge about external resources can be located in the task category. Consistent with this classification of knowledge about external resources, the four-stage model of SRL (e.g., Winne & Perry, 2000) differentiates between two broad knowledge categories: (a) knowledge about cognitive conditions (e.g., study tactics and strategies) and (b) knowledge about task conditions (e.g., instructional cues, time, and social context). In the four-stage model, knowledge about external resources is located in the “task conditions” category.

With respect to knowledge about external resources, yet another distinction must be made. Learners can have (or lack) declarative, procedural and conditional knowledge about external resources. Learners’ declarative knowledge about external resources refers to knowledge about which resources are available in a certain context or environment. Learners’ procedural knowledge about external resources refers to knowledge about how to use available resources. Conditional knowledge about external resources refers to knowledge about conditions for effective tool use (i.e., when and why to use a specific tool). Although extremely important, conditional knowledge is often overlooked. Learners are usually informed about which tools are available (declarative), and they are instructed (or they can read instructions) on how to use available tools (procedural). Often enough, however, learners are unaware of which situation is appropriate for using which type of tool (conditional) or when to refer to which (type of) external representation (e.g., a diagram).

In theories of metacognition (e.g., Efklides, 2008; Veenman, Van Hout-Wolters, & Afflerbach, 2006), conditional knowledge is basically defined as knowledge about “when” and “why to use a cognitive strategy. Such conditional knowledge is often regarded as an integral part of strategy knowledge (Paris, Cross, & Lipson, 1984) because to select and apply a strategy timely and adequately, it is essential to be able to relate strategies to specific relevant situations. It is worth noting that conditional knowledge in this sense is not the same as conditionalized knowledge in production system models such as the ACT-R theory (e.g., Anderson & Lebiere, 1998). Whereas (metacognitive) conditional knowledge is knowledge about relationships between (types of) situations and the to-be-applied knowledge or strategies, conditionalized knowledge is procedural knowledge with built-in, close connections to specific situations by internalized production rules (Renkl, Mandl, & Gruber, 1996). For purposes of simplicity, I will use the term “conditional knowledge” below when referring to metacognitive conditional knowledge.

Concerning the use of external sources, conditional knowledge could refer for example to knowledge about situations in which an online glossary should be consulted. Although most students nowadays might know how to use an online-glossary (i.e., browsing, finding, selecting and extracting information from the glossary), many may still lack knowledge about when to choose the glossary from among the many other available help facilities. The provision of conditional knowledge related to the use of help facilities should therefore enable learners to better associate outcomes of their monitoring of processing difficulties with control processes (i.e., the selection of appropriate tools), an assumption supported, for example, by an experiment of Schwonke et al. (2013). Learners provided with conditional knowledge about different help tools (e.g., a glossary, general and specific online-hints) in an intelligent tutoring environment used those tools more systematically and more strategically than learners without such support.

Noticeably, conditional knowledge about external sources does not necessarily develop just by being exposed to the resources–especially not when a CBLE is complex and the to-be-learned content is demanding. This assumption is supported by evidence that even learners with a lot of CBLE experience can display inadequate use of information sources. For example, Aleven and Koedinger (2000) found suboptimal help use in Cognitive Tutors even in learners...
who had already spent more than 500 minutes in that learning environment. Therefore, it seems necessary to explicitly support students in developing conditional knowledge about information resources so that they can make better use of available resources.

Conditional knowledge, moreover, does not necessarily grow together with strategy knowledge (Schraw, 1998). For example, learners of mathematics can usually correctly apply mathematical procedures as long as they are told which procedure to apply; they know how to apply the procedure. However, those same learners may fail when they must decide on their own which of several different procedures to use. In other words, they often do not know when to use a procedure.

Learners differ both in their knowledge about external resources as well as in their conceptions about the roles or functions of these resources and their educational usefulness (Davis, 1989). These differences are related to more general inter-individual differences about the nature of knowledge, knowing, and learning (i.e., epistemological beliefs; e.g., Perry, 1970; Hofer & Pintrich, 2002). Such conceptions may be accurate, diffuse, wrong, or even absent. In addition, such conceptions may be congruent or incongruent to the instructional designers’ intentions (Winne, 1982). Based on the CLT premise that knowledge and beliefs as represented in learners’ long-term memory form the informational basis of the central executive component of working memory (e.g., Sweller et al., 2011), it is reasonable to assume that learners’ potentially inaccurate knowledge and beliefs about such context variables can affect their cognitive, metacognitive, and self-regulatory activities. For example, Schonke, Berthold, & Renkl (2009) analyzed effects of informing students about the didactic function of diagrams in worked-out examples on learning outcomes (in the domain of mathematics). The instruction enabled low-prior knowledge students to study the worked examples more efficiently. “Informed” low-prior knowledge students paid comparatively less attention to diagrams and equations but performed better in a post-test than “uninformed” low-prior knowledge students. On the other hand, the instruction prevented high-prior knowledge students from paying too little attention to the external representations. For the latter, paying greater attention to the representations was associated with better learning outcomes. Overall, these results stress the importance of learners’ conceptions about didactic functions of external resources. More specifically, the results highlight potential relationships between learners’ conceptions and important learning process variables (here, the distribution of learners’ attention). In addition, the moderating effect of prior knowledge suggests that such conceptions can be related to learners’ prerequisites.

Taken together, most CBLEs provide learners with a certain degree of freedom. To make effective use thereof, they have to rely on relevant self-regulatory skills and general metacognitive knowledge about CLBEs, the various features of a specific environment (e.g., conditional knowledge), and the didactic conceptions behind those features. A number of empirical results support the assumption that learners often lack these self-regulatory skills and relevant knowledge, or that they possess incorrect knowledge or conceptions that are inconsistent with the instructional designers’ concepts. In light of inconsistent knowledge and skills on the part of learners, it seems reasonable to assume that CBLEs pose heavy metacognitive and self-regulatory demands on learners. In the next section, I therefore, propose considering ways to incorporate such demands into resource-oriented theories such as CTML and CLT. Exemplarily for CLT, I propose considering a type of metacognitive load (as Valeke already did in 2002 in his commentary to the special issue “Cognitive Load: Updating the theory?”).

**Extending the resource-oriented perspective**

Against this background, it might not suffice for resource-oriented theories to concentrate on intrinsic and extraneous demands on schema construction, or to concentrate primarily on core cognitive processes such as selection, organization, and integration. Rather, resource-oriented theories (such as CLT and CTML) might profit from recognizing (a) that metacognitive and self-regulatory processes exist, (b) that these processes are essential for learning, and (c) that related activities can consume working memory capacity.

In his commentary to the special issue “Cognitive load: updating the theory?” Valcke (2002) proposed extending CLT by the *Metacognitive Load* concept. In his comment, he placed special emphasis on monitoring activities (e.g., monitoring the selection and organization of sensory information or monitoring the back-and-forth storage of information from long-term memory to short-term memory). Valcke further noted that metacognitive knowledge had not been explicitly considered in any of the state-of-the-art contributions to that particular issue. He attributed this lack to the intense focus on declarative and procedural knowledge. Valcke further proposed conceptualizing metacognitive load as a part of germane load. He argued that learners besides the effort they invest in constructing
and storing schemata – also expend effort monitoring these activities. He proposed the term *germane metacognitive load* to describe this type of load. However, Valcke’s propositions have not been taken up by the CLT research community.

Nevertheless, I find the notion of working-memory load via metacognitive demands to be highly plausible and useful. We can assume that metacognitive load can be affected by the learning task, learning environment, and by learners’ prior domain-specific knowledge as well as by their general metacognitive knowledge (e.g., knowledge about learning strategies). As such, metacognitive load can be considered a potential moderating variable to explain variation in learning outcomes as the result of a specific instructional design. I further assume that metacognitive load is directly related to learning activities and learners’ interactions with the learning environment. As such, it can also be regarded as a potential mediating variable to explain effects of different instructional designs on learning outcomes.

In contrast to Valcke (2002), who conceptualized metacognitive load as a part of germane load, I rather propose conceptualizing metacognitive load as a form of working memory load that can add to either intrinsic (or germane) load or to extraneous load. In those cases where metacognitive and self-regulatory activities (such as planning, monitoring, and choosing or changing strategies and tactics) contribute to learning, these activities can be said to impose intrinsic (or germane) meta-cognitive load. In other cases in which working-memory load by metacognitive activities does not contribute to learning, I believe it should be called *extraneous metacognitive load* instead.

Whether learners’ metacognitive efforts are intrinsic (i.e., beneficial) or extraneous (i.e., detrimental) might strongly depend on learner variables as well as context factors. For example, insufficient prior domain-specific knowledge can hinder effective monitoring. In addition, sub-optimally designed learning tasks and/or learning environments (e.g., split attention formats, redundant or irrelevant information) may require unnecessary monitoring and regulatory activities. We can further assume that different learning domains (e.g., mathematics, language learning) and, more specifically, different learning tasks (solving mathematics problems; vocabulary learning) not only differ in their degree of element interactivity (i.e., intrinsic load), but also in their metacognitive and self-regulatory demands. Such differences can refer to the quantity and quality of monitoring (e.g., setting of intermediate goals) and self-regulatory activities (e.g., tactics to be chosen). For example, from a cognitive load perspective, we would argue that learning a second language’s grammar is harder than learning vocabulary because of greater element interactivity (in the presence of a degree of prior knowledge). From a metacognitive perspective, we could however argue that planning, monitoring, evaluating, and regulating activities will also differ in conjunction with these two learning tasks, and that each of these activities will probably be more demanding while mastering grammar than when learning vocabulary. One can, for instance, easily imagine that detecting an error or a misconception as well as fixing an error or overcoming a misconception is more demanding when learning grammar than when learning vocabulary. Thus we can assume that domain-specific prior knowledge and metacognitive knowledge prerequisites are directly related to an intrinsic meta-cognitive load.

**Conclusion**

Based on results from research on learning from multiple external representations and research on tool use, I have argued that there are metacognitive and self-regulation demands beyond the cognitive demands of CBL. These demands arise as a result of interaction with the external representations of the learning contents and learning task as well as interaction with a CBLE’s tools. How much these additional metacognitive and self-regulation demands tax the learners’ mental resources (especially their working memory) will largely depend on their level of metacognitive knowledge and skills. Therefore I propose considering a metacognitive type of load in resource-oriented theories such as CLT. This metacognitive load can be conceived as a type of working memory load by metacognitive or self-regulatory demands as a result of working on a learning task while interacting with a computer-based learning environment. Furthermore, metacognitive load can probably not be considered to always further learning. Rather, it may either be beneficial – when metacognitive and self-regulatory activities contribute to schema construction, ineffective – when the activities do not contribute to schema construction, or even detrimental – when the activities hinder schema construction. The quality of metacognitive demands will depend on a fit among at least three factors: (1) the learning task’s design, (2) the design of the CBLE, (3) the learners’ prerequisites (i.e., domain-specific knowledge, metacognitive knowledge and self-regulation skills). As a consequence, at least parts of the working-memory load that has so far been attributed to either intrinsic, germane, or extraneous demands on cognitive
processes (i.e., selection, organization, and integration) by the learning content, learning task, or learning environment should probably instead be attributed to demands on meta-cognitive and self-regulation processes such as planning, monitoring, or regulating. In consequence, the incorporation of a metacognitive and self-regulation-induced type of load would allow resource-oriented theories to explain a wider variety of phenomena, help these theories to represent human information processing and learning more accurately, and, as a consequence, become even more valuable for present and future instructional designers.

References


Schwonke, R., Berthold, K., & Renkl, R. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology, 23*, 1227–1243. doi:10.1002/acp.1526


Zimmerman, B. J. (1986). Becoming a self-regulated learner; Which are the key subprocesses?. *Contemporary Educational Psychology, 11*, 307–313.

The Effects of Rapid Assessments and Adaptive Restudy Prompts in Multimedia Learning

Alexander Renkl*, Irene T. Skuballa, Rolf Schwonke, Nora Harr and Jasmin Leber
Department of Psychology, University of Freiburg (Germany), Freiburg, Germany // renkl@psychologie.uni-freiburg.de // irene.skuballa@uni-tuebingen.de // schwonke@psychologie.uni-freiburg.de // harr@psychologie.uni-freiburg.de // jasmin.leber@psychologie.uni-freiburg.de

*Corresponding author

ABSTRACT
We investigated the effects of rapid assessment tasks and different adaptive restudy prompts in multimedia learning. The adaptivity was based on rapid assessment tasks that were interspersed throughout a multimedia learning environment. In Experiment 1 (N = 52 university students), we analyzed to which extent rapid assessment tasks were reactive (i.e., whether these tasks change what should be measured) and, thus, had per se positive effects on learning. In Experiment 2 (N = 41 university students), we analyzed the advantages and disadvantages of specific and unspecific restudy prompts (i.e., focus on a very specific piece of knowledge or on the corresponding knowledge sub-area). We found no reactivity associated with the assessment tasks. Most specific knowledge gaps could be closed by either type of prompt. However, unspecific prompts fostered the overall learning outcomes more than specific prompts. The present adaptation procedure is a good starting point for developing powerful adaptation mechanisms.

Keywords
Rapid assessment, Adaptive prompts, Multimedia learning, Specificity of prompts

Introduction
Educational technology provides us with learning environments which adapt learning paths to the individual's needs. Students learn by problem solving via a number of well-established systems. Their corresponding attempts can be used to adapt hints and learning tasks to their individual needs (e.g., Cognitive Tutor; http://www.carnegielearning.com; Koedinger & Corbett, 2006). Other environments provide expository instruction and present multimedia contents to learners. In these cases, explicit probes or diagnostic tasks must be interspersed for adaptation purposes (e.g., SmartBook; http://www.engadget.com/2013/01/08/mcgraw-hill-smartbook). An efficient means of such diagnosis is the use of rapid assessment procedures as developed by Kalyuga and colleagues within the framework of cognitive load theory (e.g., Kalyuga, 2008; Kalyuga & Sweller, 2005). Rapid assessment tasks can be used to detect learners’ knowledge gaps and to adapt further learning paths to these deficits.

We conducted two experiments on the effects of rapid assessment tasks and two types of adaptive restudy prompts in multimedia learning. Rapid assessments are tasks that should be fulfilled quickly and that are interspersed throughout a learning environment. The first experiment addressed the extent to which rapid assessment tasks can be regarded as a non-reactive diagnostic method, meaning that they do not change or influence what they measure (i.e., knowledge states). Note that in the first experiment, rapid assessment was not used to adapt instruction; we tested just its potential reactivity. In the second experiment, rapid assessment was used to diagnose knowledge gaps and adapt instruction. Specifically, we tested the effects of two types of restudy prompts triggered by wrong responses to rapid assessment tasks.

Adaptive learning systems
If you confront different learners with a learning environment, they are sure to differ in their learning outcomes (e.g., Ackerman & Lohman, 2006). These differences can be attributed to the learners’ varying prerequisites and the fact that a one-size-fits-all environment cannot be optimal for different learning prerequisites. One remedy is to use adaptive learning environments (e.g., Shute & Zapato-Rivera, 2008; Vandewaetere, Desmet, & Clarebout, 2011). Adaptation can refer to different sizes of grain in this context. At its coarsest, macro-adaptation refers to assigning different learning environments to different learners (Park & Lee, 2003). When the grain is fine, micro-adaptation refers to adapting instructional events during learning to a learner’s cognitive or affective states. We focused on micro-adaptation in our experiments.
Adaptive systems can react to different learner characteristics such as (prior) knowledge states, working memory capacity, cognitive styles, motivation, or emotional states (see, e.g., Vandewaetere et al., 2011). Knowledge-related variables have most frequently been used to adapt instruction (e.g., prior knowledge, knowledge states, or identified knowledge gaps). This emphasis on knowledge-related factors is not surprising, given their conceptual affinity with the knowledge-related learning goals of most adaptive systems. Furthermore, knowledge prerequisites are the most important factor for further learning both positively and negatively. Correct prior knowledge is usually the most important factor facilitating further learning (e.g., Dochy, de Rijdt, & Dyck, 2002; Kalyuga, 2012). Incorrect knowledge (e.g., misconceptions or misunderstanding) is usually the most substantial barrier for further learning (Ambrose & Lovett, 2014). In addition, research on aptitude-treatment interactions and on the expertise-reversal effect (i.e., sensible instructional features for novices lose their effectiveness with more knowledgeable learners) has clearly shown that different instructional procedures are best used for learners with different knowledge states (for an overview see Lee & Kalyuga, 2014). It goes without saying that knowledge-related indicators have often been used in adaptive systems.

Rapid assessment

Within a cognitive load framework, Kalyuga and colleagues (e.g., Kalyuga & Renkl, 2010) have produced many findings on the expertise-reversal effect. This effect means that learners with different (prior) knowledge levels profit from different instructional features (see Lee & Kalyuga, 2014). In this context, Kalyuga and colleagues (e.g., Kalyuga, 2006, 2008; Kalyuga & Sweller, 2005) developed a rapid assessment method that enables the online-diagnosis of knowledge states and deficits. For this purpose, (small) tasks, which should be answered rapidly, are interspersed throughout a learning environment. Some formats of rapid assessment tasks can be sensibly applied just in procedural domains (e.g., the first-step method that requires learners to provide a first solution step to a problem). A format applicable in most domains is rapid verification. Rapid verification tasks present solution steps or statements, and the learners must quickly determine whether the presented information is right or wrong (see also Roelle, Berthold, & Renkl, 2014).

Kalyuga and Sweller (2004) investigated the effects of adapting instruction on the basis of rapid assessment. They exploited the finding that with a learner’s growing knowledge level, subsequent learning tasks are best for acquiring cognitive skills: first fully-worked examples, then faded-worked examples (i.e., worked examples containing some gaps to be filled in), and finally problem solving tasks (e.g., Renkl, 2014; Renkl & Atkinson, 2003). In the adaptation condition, learners studied fully-worked examples, faded-worked examples, or problem solving exercises depending on their prior knowledge level and their progress during the learning phase, as assessed by rapid assessment tasks. In the control group, each learner followed the learning path of a yoked learner from the adaptation group. There were no rapid assessment tasks in the control group to monitor learners’ progress during the learning phase. Kalyuga and Sweller (2004) observed superior learning outcomes in the adaptation group. However, it is not totally clear to what extent this advantage was due only to the adaptive presentation of learning tasks or to working on the additional test tasks as well. The latter activity can enhance learning outcomes, as research on the testing effect has revealed (Rawson & Dunlosky, 2012; Roediger, Putnam, & Smith, 2011). Against this background, it would be interesting to test whether just working on rapid assessment tasks already exerts positive effects on learning outcomes.

Study prompts

Prompts are questions or hints provided to learners. They are designed to induce productive processing of learning materials (e.g., Devolder, van Braak, & Tondeur, 2012; Pressley et al., 1992). The type of processing that is induced depends on its specific purpose. For example, prompts were successfully employed to foster self-explanations (Chi, de Leeuw, Chiu, & LaVancher, 1994), comparison of cases (Gentner, Loewenstein, & Thompson, 2003), or self-regulation activities (Bannert & Reimann, 2012). In some learning environments, the prompts are adaptive in the sense that they depend on the learners’ prior behavior (e.g., Graesser, Jeon, & Dufty, 2008; Nückles, Hübner, Dümér, & Renkl, 2010).

Although there are several studies comparing different types of prompts (e.g., Ifenthaler, 2012; Nückles, Hübner, & Renkl, 2009), there is hardly any evidence as to which type of prompts are best used for closing knowledge gaps (see also Devolder et al., 2012). In some cases, learners’ incorrect answers might indicate that a very specific piece of
knowledge is missing (e.g., the fact that it is the nucleus where DNA doubles during mitosis). In other cases, the corresponding sub-area of the learning contents might be missing as well (e.g., what happens in general in the nucleus during mitosis). If a prompt just encourages learners to close the very specific knowledge gap, the potentially underlying problem (that the entire sub-area is unknown) goes unaddressed. In contrast, if a prompt encourages not just looking up the specific missing piece of knowledge, but considering the knowledge sub-area as well, broader effects on learning could be expected. On the other hand, unspecific prompts might be less efficient than specific prompts when just such a specific knowledge gap should be closed. In addition, unspecific prompts might induce unnecessary processing of already understood materials (see the redundancy effect in cognitive load theory; Sweller, Ayres, & Kalyuga, 2011).

Overview of the present experiments

We conducted two experiments. Each investigated one main element of our adaptation procedure (i.e., rapid assessment-based provision of restudy prompts). In Experiment 1, we analyzed to which extent rapid assessment tasks have per se positive effects on learning outcomes. In addition, we considered factors that might contribute to such an effect (e.g., rapid assessment tasks might motivate deeper processing on the following materials). In Experiment 2, we analyzed the advantages and disadvantages of different types of restudy prompts that varied in their focus (i.e., on a very specific piece of knowledge or on the corresponding knowledge sub-area). The specific hypotheses tested in these two experiments are presented right before the corresponding method sections.

Experiment 1

We investigated whether adding diagnostic tasks to a learning environment is reactive in the sense that these tasks “alone” already foster knowledge acquisition. More specifically, we tested the potential effect of rapid assessment tasks on learning outcomes (Learning-Outcomes Hypothesis). Findings on the testing effect suggest that there might be such an effect (e.g., Rawson & Dunlosky, 2012; Roediger et al., 2011). As the assessment tasks also familiarize students with test tasks from the learning domain, we expected that the learners with rapid assessment tasks would perceive the problems in the test on learning outcomes (posttest) as easier and requiring less mental effort (Subjective-Load Hypothesis).

In addition, we explored a number of potential mediational mechanisms that may account for a learning effect: The rapid assessment tasks may orient the learners about the learning goals (see Roediger et al., 2011) (Orientation Hypothesis). The expectation of up-coming test tasks may motivate longer (Learning-Time Hypothesis) and deeper processing of the materials (Roediger et al., 2011) (Reflection Hypothesis). Providing learners with the opportunity to check what they have learned may make them perceive the learning environment as more interesting and useful. In addition, if rapid assessment tasks help learners comprehend the learning contents, this factor would also contribute to situational interest (Schraw, Flowerday, & Lehmann, 2001) (Interest Hypothesis). We planned to test mediation effects in the case of significant effects on the variables that might explain a possible effect of rapid assessment on learning outcomes (e.g., orientation about learning goals or learning time).

Method

Sample and design

Fifty-two university students from a psychology program took part in this study (age: $M = 24.63$, $SD = 5.65$). They received study credits for participation. The students were randomly assigned to two conditions ($n = 26$ in each condition): computer-based learning environment about mitosis with or without rapid assessment tasks (in form of rapid verification). No feedback was provided for the assessment tasks as we wanted to test the reactivity of the tasks themselves. Note also that the assessment tasks were not used for adaptation. The central dependent measure referred to the learning outcomes and the subjective load while working on the posttest. In addition, potential mediators were assessed: orientation about the learning goals, learning time, reflection, and situational interest.
**Learning environment**

The learning environment topic was the process of mitosis. The contents were provided by text and pictures (Figure 1). All participants were instructed to study the learning program at their own pace. They should prepare for a final test (posttest) that was “epitomized” by two exemplary items.

**Figure 1.** Screenshot from the multimedia learning environment on mitosis (Experiment 1)

The learning environment consisted of (a) an introduction including a general overview of the cell structure and of the different mitotic phases and (b) three blocks with two or three mitotic phases. The phases were explained and illustrated by schematic and realistic pictures. In order to emphasize the changes over the different stages of mitosis, the single pages simultaneously provided information about subsequent phases (Figure 1). After each block, the learners rated how well informed they felt about the learning goals, how interested they were, how much they reflected about the interrelations between the presented information, and how much they tried to anticipate what comes next (scale 1 to 7; 1: not at all).

In the rapid-assessment group, we presented 12 rapid verification tasks after each of the three blocks (4 tasks on cell structures, 4 tasks on processes, 4 tasks on functions; Kalyuga, Renkl, & Paas, 2010). Thus, there were 36 assessment tasks all in all. An exemplary verification task contained the following statement: “The microtubules form the mitotic spindle.” The learners had to click the “right” or “wrong” button. Half of the 36 statements were wrong. After 15 seconds the task disappeared. After 4 rapid assessment tasks, there was a question on how difficult it was to answer these tasks. However, we did not analyze data that were only available in the condition with rapid assessment further.

Learners from the rapid-assessment group worked on the diagnostic tasks after each block. Learners from the group without rapid assessment did not work on any tasks referring to the presented contents; there was no substitute for the rapid assessment tasks in this group.

**Instruments**

A pretest on prior knowledge about mitosis should primarily check the comparability of the experimental groups. It consisted of nine items (e.g., “Please write down five parts of the human cell”). For these open items, we provided 1 point for each aspect that was included in expert answers. Two raters coded the answers from 11 participants (i.e., about 20%). These raters reached high agreement, indicated by an ICC of .933; disagreements were resolved by discussion. We obtained an estimate of .77 (Cronbach’s alpha).
We assessed reflection with two items (i.e., how much they reflected about the interrelations within the presented information and how hard they tried to anticipate what comes next) that were present three times (i.e., after each rapid assessment block). As both sub-scales were highly correlated ($r = .68$, $p < .001$) we aggregated them to an overall reflection score (Cronbach’s alpha: .90). Situational interest was assessed with one item (“This learning block on mitosis was interesting”) that was presented three times (Cronbach’s alpha: .92). Finally, one item assessed three times the extent to which the learners felt oriented about the learning goals (“I have an idea about what I should learn”; Cronbach’s alpha: .88).

The posttest consisted of 22 items which were primarily verbal (e.g., “Please describe what happens during cytokinesis”) or primarily pictorial (e.g., question about a schematic picture: “What is wrong in this schematic picture?”). Most of these items (i.e., 16) required open answers. We provided 1 point for each aspect that was included in expert answers. Two raters coded the answers from 11 participants (i.e., about 20%). These raters reached high agreement, indicated by an ICC of .969; disagreements were resolved by discussion. We obtained an estimate of internal consistency of .83 (Cronbach’s alpha) for the overall posttest score (including 22 items). After each posttest item, we asked the learners to indicate on a 10-point rating scale how difficult the problem was and how much mental effort they had invested. We aggregated these ratings over the different posttest items and obtained a reliability estimate of .93 (Cronbach’s alpha) for subjective difficulty and of .96 (Cronbach’s alpha) for mental effort. Subjective difficulty and mental effort are both often used as measures of cognitive load (see, e.g., van Gog & Paas, 2008). However, as both measures correlated just moderately ($r = .36$, $p = .009$) we treated them separately in the following analyses (see also van Gog & Paas, 2008).

Procedure

The experimental sessions took place in small groups containing 5 to 12 participants. The students were welcomed in a computer room and instructed to press the “Start” button on their screens. The entire session was implemented on the computer. First, the students filled in a questionnaire on demographic data and worked on the pretest. Then, they received a short overview of the learning contents and two exemplary test questions that they should be able to answer after studying the learning environment. In the following, the procedure differed in certain respects between conditions. The students without rapid assessment proceeded directly to the learning environment. The students with rapid assessment received a short practice phase showing how to work with rapid assessment tasks. This phase was designed to prepare them for reacting “rapidly” even to the first assessment tasks. Afterwards, the students in the rapid-assessment condition worked on the same learning environment as the group without rapid assessment, except that there were three interspersed blocks with rapid assessment tasks. Note, however, that the learning phase was divided in three blocks for all learners because after each block we assessed interest, orientation about learning goals, and reflection. At the end of the session, all participants worked on the posttest. Finally, they received their credits for participation.

Results

For all analyses, we used an alpha level of .05. We report $d$ as effect-size measure that was interpreted as follows: $d = .20$ as small effect, $d = .50$ as medium effect, and $d > .80$ as large effect (Cohen, 1988). Table 1 shows the means and standards deviations of the central variables in both conditions. Due to some missing data the sample size for different analyses varied between 50 and 52 participants.

The learners in the two conditions did not differ significantly in their prior knowledge, $t(50) = 0.54$, $p = .589$, $d = 0.15$. As the possible maximum pretest score was 38, the overall mean of 8.10 (about 23% correct) indicated a low level of prior knowledge. There were no significant differences between groups with respect to age, semesters at university, gender, German as mother tongue, or prior biology courses (all $p$s > .20). Hence, the learners in both conditions had comparable learning prerequisites.

Overall, there were 36 rapid assessment tasks. On average 24.69 ($SD = 4.19$) tasks were correctly answered. Hence, they were far from trivial but also easy enough to be answered correctly in two-thirds of the cases.
Table 1. Means (Standard deviations in brackets) of important variables for both conditions (Experiment 1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>No rapid assessment (n = 26)</th>
<th>Rapid assessment (n = 26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge (no. correct)</td>
<td>8.46 (4.19)</td>
<td>9.15 (4.95)</td>
</tr>
<tr>
<td>Learning outcomes (no. correct)</td>
<td>37.78 (10.86)</td>
<td>34.36 (12.05)</td>
</tr>
<tr>
<td>Mental effort (posttest scale from 1 to 7)</td>
<td>5.82 (1.63)</td>
<td>4.91 (1.80)</td>
</tr>
<tr>
<td>Difficulty (posttest; 1 to 7)</td>
<td>6.15 (1.43)</td>
<td>5.74 (1.99)</td>
</tr>
<tr>
<td>Learning time (min.)</td>
<td>11.78 (3.46)</td>
<td>11.21 (3.81)</td>
</tr>
<tr>
<td>Orientation about learning goals (1 to 7)</td>
<td>5.46 (1.17)</td>
<td>5.04 (1.30)</td>
</tr>
<tr>
<td>Reflection (1 to 7)</td>
<td>4.29 (1.24)</td>
<td>3.92 (1.35)</td>
</tr>
<tr>
<td>Situational interest (1 to 7)</td>
<td>4.52 (1.52)</td>
<td>4.17 (1.66)</td>
</tr>
</tbody>
</table>

We detected no significant group differences with respect to learning outcomes, \( t(48) = -1.54, p = .297, d = -.30 \) (Table 1). For exploratory reasons, we also checked whether the learning effect was moderated by prior knowledge (ATI effect). However, the respective interaction term did not reach the level of statistical significance, \( F(1,46) = 1.09, p = .302 \). Based on these findings, we rejected the Learning-Outcomes Hypothesis.

We found no significant group differences with respect to the perceived difficulty of the posttest, \( t(49) = -0.85, p = .401, d = -.24 \), or to mental effort during the posttest, \( t(50) = -1.92, p = .061, d = -.53 \). The Subjective-Load Hypothesis was not confirmed. Both groups spent on average a bit more than 11 minutes on the learning contents (without the time on assessment tasks); there was no significant group difference. \( t(50) = -0.57, p = .574, d = -.16 \). Hence, we rejected the Learning-Time Hypothesis. Note, however, that the rapid-assessment group spent an additional 3.12 minutes (\( SD = 0.93 \)) on the rapid assessment tasks.

There were no significant group differences with respect to the learners’ perceived orientation about learning goals, \( t(50) = -1.23, p = .223, d = -.35 \), to reflections on the learning contents, \( t(50) = -1.01, p = .315, d = -.29 \), or to situational interest, \( t(50) = -0.81, p = .421, d = -.22 \). Hence, we rejected the Orientation Hypothesis, the Reflection Hypothesis, and the Interest Hypothesis.

One research question referred to potential mediation effects. Rapid assessment might have fostered learning outcomes via more reflection, time on the learning contents, situational interest, or better orientation about learning objectives. However, as we identified no significant group differences concerning the latter variables, they cannot be regarded as mediators. Nevertheless, the relations between the potential mediators and learning outcomes should be reported: \( r = .39, p = .006 \), for learning time; \( r = .36, p = .011 \), for orientation about learning goals; \( r = .39, p = .005 \), for reflection; \( r = .46, p = .001 \), for situational interest. These significant correlations can be tentatively interpreted as evidence that the assessed variables were relevant for learning in the present context.

**Discussion**

The hypotheses on potential effects of rapid assessment were not confirmed. This is good news in the present case, as we wanted to test whether rapid assessment could be regarded as a non-reactive – or just minimally reactive – diagnostic procedure. Against the background of our results, rapid assessment does not seem to be reactive.

A possible objection against interpreting the present findings as indicating the absence of reactivity might be that the reactivity effects might have just failed to reach the level of statistical significance (e.g., because of a lack of statistical power). Note, however, that there were not only “missing” significant effects, but also descriptive mean differences that in most cases were in favor of the group without rapid assessment (Table 1). Only the perceived difficulty and mental effort with regard to the posttest were descriptively lower in the rapid-assessment condition. Hence, there was hardly any indication for reactivity. Another alternative explanation for the missing effects could be that the testing dose was too low. However, this explanation is unlikely to be true because there were 36 rapid verification tasks. Hence, the effects of adaptation via rapid assessment reported in the literature (e.g., Kalyuga & Sweller, 2004, 2005) are probably genuine adaptation effects and not just effects of the additional rapid assessment tasks.
As rapid assessment was revealed as a non-reactive assessment procedure, we used this method in a further study to identify knowledge gaps. In the case of knowledge gaps, the learners restudied the contents which they had not learned well.

**Experiment 2**

Incorrect answers to rapid assessment tasks might indicate that a very specific piece of knowledge is missing or that there are deficits with respect to the corresponding sub-area of the learning contents. A restudy prompt that focuses on the specific piece of knowledge tested in rapid assessment should be the most straightforward way to repair a very specific deficit, but it would not address broader deficits in the corresponding sub-area.

A more general restudy prompt should have the advantage of broader effects on the knowledge sub-area. Expecting such broader effects on learning when prompts pose more general questions is in line with the findings of Vollmeyer and Burns (2002). They took up the goal specificity effect, as revealed by cognitive load research (e.g., Sweller et al., 2011), and tested whether learners acquire more declarative knowledge when they pursue more general goals during learning in a hypermedia (multimedia) environment. Actually, it is more general goals that lead to better learning. Brunstein and Krems (2005) observed similar results in conjunction with learning from hypertext. Nevertheless, unspecific prompts or goals might also have disadvantages. If the prompts focus on the specific knowledge gaps identified by rapid assessment, these knowledge gaps might be closed more reliably than when the prompts ask for more general exploration of a knowledge area.

In this context, we assumed a differential effectiveness of specific prompts exclusively directing attention to a specific piece of knowledge and of unspecific prompts directing attention to the corresponding knowledge sub-area. Specific prompts should close knowledge gaps more effectively and they should foster the type of knowledge that is directly tapped by the rapid assessment tasks, namely the most important (central) knowledge about the mitotic process. The more general prompts should be more effective when knowledge is being fostered that extends beyond the central issues in the mitotic process (e.g., cell structure in general, or transfer tasks).

In summary, we addressed the following hypotheses: Specific prompts are more effective in repairing specific knowledge gaps identified by rapid assessment (Knowledge-Repair Hypothesis). Specific prompts are more effective in fostering knowledge about central issues of the mitotic process (Central-Learning-Contents Hypothesis). Unspecific prompts are more effective in fostering knowledge about more general issues related to mitosis (Broader-Learning Hypothesis). Unspecific prompts are more effective in fostering situational interest (Interest Hypothesis).

**Method**

**Sample and design**

Forty-one university students recruited mainly from a psychology department took part in this study (age: $M = 22.41$, $SD = 2.76$). They received either study credits or 15 Euros for participation. The students were randomly assigned to two conditions: Adaptive restudy with (1) specific prompts (directing attention to the specific missing piece of knowledge; $n = 20$) or (2) unspecific prompts (directing attention to the knowledge area “embedding” the specific information; $n = 21$). The dependent measures referred to the learning outcomes in terms of repaired knowledge gaps, knowledge acquisition about the mitotic process as well as about general issues related to mitosis, and to situational interest in the learning contents.
Materials

The pretest assessing prior knowledge contained 8 items. One open item also included in the posttest asked the students to describe the process of mitosis in about 150 words. We assigned 1 point for each of the 24 aspects included in an expert answer. Two raters coded the answers for about 25% of the participants (i.e., 10 persons). We observed high interrater agreement as revealed by an ICC of .931; disagreements were resolved by discussion. Given this high agreement, the rest of the answers on this pretest and posttest item were coded by a single rater. In addition, we determined the internal consistency of this pretest measure (regarding the 24 aspects as items). We obtained a satisfying Cronbach’s alpha coefficient of .79 for this measure of prior topic knowledge (on the mitotic process, i.e., the central learning contents). Seven items that had not been included in the posttest asked for other useful prior knowledge when learning about mitosis (e.g., “List five elements in the human cell.”). For the open items, we again assigned 1 point for each aspect included in expert answers. Two raters coded the answers from 10 participants (i.e., about 25%). These raters also reached high agreement as indicated by an ICC of .932; disagreements were resolved by discussion. We obtained an internal consistency estimate of .73 (Cronbach’s alpha) for this score of prerequisite knowledge. As both prior knowledge scores correlated with $r = .81$, $p < .001$, we computed a combined score for prior knowledge.

In the beginning of the posttest, the students worked again on all rapid assessment tasks that they had answered incorrectly during the learning phase. Thereby we could see to what extent prompted restudy closed the specific knowledge gaps. Learning outcomes on central learning contents (i.e., mitotic process) were assessed by the same item as in the pretest, that is, the students were requested to describe this process in about 150 words. We again determined a reliability estimate in terms of internal consistency (Cronbach’s alpha: .83). In addition, we used 16 multiple-choice items (e.g., the learners inspected a schematic picture of the mitotic process with an error in it; they had to mark one of four presumable errors) and 5 open items (e.g., three marked structures in a realistic picture had to be named). Two raters coded the answers to the five open questions for about 25% of the participants (i.e., 10 persons). As in the case of the open pretest items, the raters referred to expert answers to score the participants’ answers. We again observed high interrater agreement, as determined by an ICC of .947; disagreements were resolved by discussion. Given this high agreement, the rest of the answers were coded by a single rater. For the posttest score including the 16 multiple-choice items and the 5 open items, we determined a Cronbach’s alpha of .77. Both posttest scores correlated with $r = .59$, $p < .001$, which is substantial but does not preclude separate analysis of both scores.

Situational interest was assessed by 10 items that were to be answered on a Likert scale from 1 to 7 (1: I do not at all agree; 7: I fully agree): four items referred to the value-related component of situational interest (e.g., “The topic of the learning environment is important”) and six items to the emotional component (e.g., “I was bored while I worked on the learning environment”; negatively keyed item) (Schiefele & Krapp, 1996). We determined a reliability estimate of .93 (Cronbach’s alpha).

Learning environment and experimental variation

We took the learning environment from Experiment 1 (i.e., same content, same number of pages, same page design, etc.). However, we made some modifications to “repair” some sub-optimal features of the text and the pictures. In addition, the participants worked on the rapid assessment tasks after smaller blocks of learning contents (e.g., each mitotic phase) in this experiment. We wanted to close the knowledge gaps more or less immediately so they would not impede further learning.

Overall, the learning environment comprised ten blocks. Each block depicted information on cell division followed by three rapid assessment tasks. These tasks presented statements, and the participants had to indicate whether the statements were right or wrong. In the case of incorrect answers, the learners were automatically directed to the corresponding learning contents for restudy in order to close their knowledge gaps.

Restudying was prompted differently between conditions. The prompts guided the learners on how to process the information on the page to which the learners had been re-directed. In the specific prompts condition, the relevant passages were highlighted by darkening the less relevant information on the page (Figure 2). Note that the darkened text passages could still be read relatively easily so that the presented information did not differ between conditions.
The prompt consisted of the request to restudy the relevant passage in order to solve the task correctly, and the task was repeated (Figure 2). In the unspecific prompts condition, the learners were asked to restudy and figure out both the direct answer to the question and to explore the broader context. For example, if the rapid assessment task “The equatorial plane is a straight plate dividing the cell during the metaphase” was answered incorrectly, the prompt “Detect what the equatorial plane is” was displayed together with the page containing the relevant information (Figure 3).

**Figure 2.** Screenshot with a specific prompt (Experiment 2)

**Figure 3.** Screenshot with an unspecific prompt (Experiment 2)

**Procedure**

Participants were tested in individual sessions. The entire session was carried out on the computer. After a brief welcome, the students were familiarized with the computer system and asked to fill in a questionnaire on demographical data and to work on the pretest. Before continuing with the learning phase, participants were informed about the architecture of the learning environment, the rapid assessment tasks, and the restudy procedure.
Afterwards the students worked on the learning environment in which the experimental variation took place. After the learning environment, the students filled in the questionnaire on situational interest and worked on the posttest. At the end of these procedures, they received study credits or 15 Euros for participation.

Results

As in Experiment 1, we used an alpha level of .05, and we reported $d$ as effect-size measure that was interpreted as follows: $d = .20$ as small effect, $d = .50$ as medium effect, and $d > .80$ as large effect. Table 2 shows the means and standards deviation of the central variables in both conditions. The sample size used for the statistical analyses varied slightly between 40 and 41 because of missing data.

<table>
<thead>
<tr>
<th></th>
<th>Unspecific prompts $(n = 21)$</th>
<th>Specific prompts $(n = 20)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge (z score)</td>
<td>0.23 (1.08)</td>
<td>-0.24 (0.86)</td>
</tr>
<tr>
<td>Learning time (min.)</td>
<td>21.10 (4.41)</td>
<td>21.94 (6.93)</td>
</tr>
<tr>
<td>Rapid assessment task (no. correct)</td>
<td>24.00 (3.36)</td>
<td>22.85 (3.91)</td>
</tr>
<tr>
<td>Repaired knowledge gaps (%)</td>
<td>82.64 (21.90)</td>
<td>82.07 (15.62)</td>
</tr>
<tr>
<td>Knowledge of mitotic process (no. correct)</td>
<td>8.81 (4.45)</td>
<td>5.95 (4.34)</td>
</tr>
<tr>
<td>General knowledge (no. correct)</td>
<td>25.88 (5.42)</td>
<td>22.30 (5.65)</td>
</tr>
<tr>
<td>Situational interest (scale from 1 to 7)</td>
<td>5.72 (0.73)</td>
<td>4.77 (1.58)</td>
</tr>
</tbody>
</table>

The two conditions did not differ significantly in their prior knowledge (overall score), $t(39) = 1.53$, $p = .113$, $d = 0.48$. However, there was a descriptive difference between conditions of almost half a standard deviation. Hence, we planned to statistically control for prior knowledge in the following analyses by ANCOVAs. However, the pretest scores did not significantly predict any of the performance measures, so that ANCOVAs made no sense. As the participants reached on average a percentage-correct score of about 14% in the pretest, they can be regarded as learners with a low level of prior knowledge.

There were no significant group differences with respect to age, semesters at university, German as mother tongue, or prior biology courses (all $p$s > .12). The participants’ gender was not equally distributed across conditions (specific prompts: 19 females and 2 males; unspecific prompts: 11 females and 9 males; chi$^2(1) = 6.67$, $p = .010$). However, gender was not related to repaired knowledge gaps, knowledge about the mitotic process, and more general knowledge (all $p$s > .69). Female and male participants did not significantly differ in interest either, $t(38) = 1.52$, $p = .137$. Nor was there a difference between groups with respect to learning time (Table 2), $t(38) = -0.46$, $p = .648$.

There were 30 rapid assessment tasks. On average 23.43 ($SD = 3.64$) tasks were correctly answered. Hence, there were about 78% correct answers. The conditions did not differ in this respect (Table 2), $t(38) = 0.99$, $p = .325$. Overall, the initially incorrect answers to the rapid assessment tasks were corrected in slightly more than 82% of the cases. The two conditions did not differ significantly in how successfully the specific knowledge gaps were repaired (Table 2), $t(38) = 0.094$, $p = .926$. Both types of prompts seemed to be suited for closing knowledge gaps. Hence, we rejected the Knowledge-Repair Hypothesis.

With respect to learning outcomes about the mitotic process (i.e., central learning contents), we found – in contrast to our expectations – the unspecific prompts to be superior (Table 2), $t(39) = 2.08$, $p = .044$, $d = 0.65$. Hence, our Central-Learning-Contents Hypothesis had to be rejected. With regard to the more general learning outcomes, we found the unspecific prompts to be superior (Table 2), $t(39) = 2.07$, $p = .045$, $d = 0.65$, confirming the Broader-Learning Hypothesis.

Learners with unspecific prompts stated higher interest in the learning contents than those with specific prompts (Table 2), $t(26.73) = 2.45$ (test for unequal variances), $p = .021$, $d = 0.77$. We thus confirmed the Interest Hypothesis.

Post hoc we tested for exploratory reasons whether the superior learning outcomes of the group with unspecific prompts were mediated by enhanced situational interest. In fact, situational interest correlated substantially with both learning outcome measures (knowledge about the mitotic process: $r = .48$, $p = .002$; more general knowledge: $r$
These findings point to mediation effects (Hayes, 2013). When directly testing the mediation effects by a bootstrapping procedure (number of bootstrap samples = 1000) as suggested by Hayes, we found that situational interest significantly mediated the effects of unspecific (vs. specific) restudy prompts on learning outcomes in terms of general knowledge and knowledge about the mitotic process. Unspecific prompts significantly heightened interest ($b = -0.95, SE = 0.388, t = -2.448, p = .019, LCL = -1.736, UCL = -0.165$), and interest was a significant predictor of general knowledge ($b = 2.17, SE = 0.659, t = 3.306, p = .002, LCL = 0.843, UCL = 3.151$) controlling for direct effects of the specificity of prompts. At the same time, the specificity of prompts (i.e., the experimental condition) lost its predictive value ($b = 1.430, SE = 1.697, t = -0.843, p = .405, LCL = -4.868, UCL = 2.007$) for general knowledge. A similar pattern appeared for an indirect effect of prompt specificity via interest on knowledge about the mitotic process. We found interest to be a significant predictor of knowledge about the mitotic process ($b = 1.51, SE = 0.540, t = 2.801, p = .008, LCL = 0.419, UCL = 2.608$) while the specificity of prompts lost its predictive value ($b = 1.312, SE = 1.391, t = -0.944, p = .355, LCL = -4.130, UCL = 1.506$). As an effect size measure for the indirect (mediation) effect, we calculated the ratio of the indirect effect of the prompt specificity (as mediated by interest) to the total effect of the prompt specificity on learning outcomes. With respect to the general-knowledge outcome variable, the indirect effect represented 59.1% of the total effect. With respect to the knowledge-about-the-mitotic-process outcome variable, the indirect effect represented 52.3% of the total effect.

**Discussion**

We assumed differential effects of specific and unspecific prompts on learning outcomes. Specific prompts should be preferable to repair specific knowledge gaps and foster knowledge directly related to the mitotic process (Knowledge-Repair Hypothesis and Central-Learning-Contents Hypothesis). Unspecific prompts should be preferable in fostering knowledge about general issues related to mitosis (Broader-Learning Hypothesis). What we found was the general superiority of unspecific prompts, except that both types of prompts repaired the specific knowledge gaps in most cases.

In addition, our findings suggest that unspecific prompts foster situational interest. This effect is in line with self-determination theory that assumes that enhancing learners’ perceived autonomy enhances motivation (e.g., Niemiec & Ryan, 2009). However, it will be up to a future study to determine whether situational interest is actually fostered by an increased level of perceived autonomy when working with unspecific prompts. In any case, our findings from the mediation analysis indicate that the positive learning effects of unspecific prompts are (partly) due to their motivational advantage.

**Overall discussion**

In Experiment 1, we identified no indication that rapid assessment tasks are reactive with respect to learning outcomes. This finding can be considered to be good news and bad news. From a practical perspective, it is bad news because it would be terrific if the diagnostic procedure underlying an adaptation procedure already had positive learning effects. From an experimental stance, it is good news because this non-reactivity makes it easier to determine pure effects of the adaptive provision of restudy prompts.

The non-reactivity of rapid assessment on learning outcomes might be surprising in light of evidence reported in the testing-effect literature (for reviews see, e.g., Rawson & Dunlosky, 2012; Roediger et al., 2011). Note, however, that we did not provide feedback in the first study. Feedback is discussed as an important ingredient when testing procedures are designed to foster learning (e.g., Rawson & Dunlosky, 2012; Roediger et al., 2011). We are well aware that a testing effect can also occur when no feedback is provided (e.g., Roediger & Karpicke, 2006). On the other hand, some testing effect studies have found that testing without feedback can even exert negative effects (i.e., consolidating incorrect knowledge: e.g., Roediger & Marsh, 2005).

Another possible cause for the present findings’ divergence from the testing effect may be due to the fact that we administered an immediate posttest only. Testing effects are typically found in delayed posttests (e.g., Roediger & Karpicke, 2006). Hence, it would be sensible to use a delayed posttest in further studies.
Another limitation associated with Experiment 1 is that we did not test all possible mediation effects. For example, the rapid assessment tasks could have cancelled out learners’ over-confidence or illusion of knowledge which can, in turn, foster learning outcomes (e.g., Rawson & Dunlosky, 2012). However, as there was no overall effect on learning, we tentatively assume that we did not miss an important mediational mechanism.

The positive effects of unspecific prompts on learning outcomes suggest that suggest that adaptive systems should close knowledge gaps not just by addressing the very specific knowledge gaps but by more general prompts for restudy. Such unspecific restudy has both cognitive and motivational advantages. Our findings demonstrate that the present effect of goal specificity is caused not only by cognitive factors, as suggested by prior research (e.g., Vollmeyer & Burns, 2002: learning strategies and use of resources), but by motivational factors such as interest as well.

Overall, the present experiments inform about the effects of our adaptation procedure consisting of a rapid assessment-based provision of restudy prompts. Rapid assessment does not per se have a positive effect on learning outcomes. However, such diagnostic tasks can be used to close specific knowledge gaps that have been identified (i.e., remediation rate over 80%). To exert broader effects on learning outcomes, rapid assessment should be combined with unspecific restudy prompts.

Note, however, that the performance on the learning outcome measures, even in the condition with the unspecific restudy prompts, was well below the ceiling. This sub-optimal performance might be due to the fact that our pre-determined rapid assessment tasks were not able to reveal all or at least most of the knowledge deficits that the learners had during learning. A potential remedy might be to use online data such as eye movements to collect “suspicious facts” (e.g., very little time on specific information or regressions to already-studied materials). Such indicators could be used to adaptively select rapid assessment tasks that could then verify or falsify the “suspicion,” leading to the corresponding decision to present or skip a restudy prompt. Such a “double-adaptive” system (adaptive selection of rapid assessment tasks and adaptive selection of prompts) might help to further optimize learning.

Overall, we provide evidence that knowledge gaps during learning can be detected and closed by a rapid assessment-based adaptation procedure that presents (unspecific) restudy prompts. This procedure should continue to be improved in order to detect knowledge gaps during learning more sensitively. With the present two experiments, we have laid a sound foundation for future work on optimizing an adaptation procedure for expository multimedia learning environments.

References


Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A General role of analogical encoding. Journal of Educational Psychology, 95, 393-408.


Using Cognitive Load Theory to Tailor Instruction to Levels of Accounting Students’ Expertise

Paul Blayney¹*, Slava Kalyuga² and John Sweller²

¹School of Business, University of Sydney, Australia // ²School of Education, University of New South Wales, Australia // paul.blayney@sydney.edu.au // s.kalyuga@unsw.edu.au // j.sweller@unsw.edu.au

*Corresponding author

ABSTRACT
Tailoring of instructional methods to learner levels of expertise may reduce extraneous cognitive load and improve learning. Contemporary technology-based learning environments have the potential to substantially enable learner-adapted instruction. This paper investigates the effects of adaptive instruction based on using the isolated-interactive elements effect that occurs when learners who are initially presented with elements of information in an isolated, non-interactive form (followed by a fully interactive form) outperform those who are presented with the same information only in a fully interactive form. Cognitive load theory explains the effect for novice learners by their potential cognitive overload when dealing with a fully interactive form of instruction from the beginning. However, according to the expertise reversal effect in cognitive load theory, the effect may reverse for relatively more knowledgeable learners. Experiment 1 found that more knowledgeable accounting students performed better with interactive rather than isolated presentations. For less knowledgeable learners, there was no statistically significant performance difference between the presentation formats. Thus, there was a significant interaction between the instructional procedures and levels of learner prior knowledge as an indicator of an expertise reversal effect. In one of the two conditions used in Experiment 2, information was adaptively presented in isolated form to less knowledgeable learners but in interactive form to more knowledgeable learners (based on the pre-tests of learner prior knowledge). In another (control) group, students were randomly allocated to isolated and interactive instructional formats irrespective of levels of their prior knowledge. As expected, the adaptive instruction group was superior to the non-adaptive group. The paper concludes with implications for the technology enabled design of learner-tailored instructional presentations.

Keywords
Adaptive instruction, Cognitive load theory, Isolated-interacting elements effect, Expertise reversal effect, Accounting training

Introduction
Cognitive load theory has been extensively used to investigate the implications of human cognitive architecture for instruction and learning in different domains (Sweller, Ayres, & Kalyuga, 2011). This instructional theory draws on various characteristics of the major components of human cognitive architecture, primarily working memory as a conscious information processor and long-term memory as our knowledge base, to advance teaching and learning techniques. The theory recognises the limited capacity of working memory (Miller, 1956; Baddeley, 1992) when dealing with novel information as well as the critical role of available knowledge structures in long-term memory for learning and performance (De Groot, 1965; Chase & Simon, 1973). Together, these two factors determine the magnitude of working memory load which is essentially cognitive load.

Tailoring the design of instructional procedures and formats to levels of learner prior knowledge is one of the essential recommendations of this theory based on research on the expertise reversal effect (Kalyuga, 2007). This research has demonstrated that different instructional methods are suitable for learners with different levels of expertise in a task domain. Therefore, the cognitive load consequences of using various instructional methods can be optimized if these methods are intentionally tailored to individual learners’ levels of expertise (learner-tailored or adaptive instruction). Any type of effective practical implementation of individualized, adaptive, learner-tailored instruction today requires the use of modern technology. Without the use of technology, adaptive instruction methods have usually been limited to individual or small group face to face teaching. The experiments in this study rely on the most widely used technology in the accounting profession – the Excel spreadsheet. This study’s adaptive alteration of instructional methods on a large scale basis is substantially enabled by spreadsheet technology-based learning environments.
The expertise reversal effect

The expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, 2007) recognises the differential effectiveness of instructional methods and techniques according to the level of a person’s expertise or prior knowledge. This effect provides a theoretical basis for suggesting that less knowledgeable learners (e.g., novices) should be provided with a high degree of instructional support for their learning activities. On the other hand, relatively more knowledgeable learners (e.g., experts) should be provided with a lower level of assistance thus allowing them to use their previously developed schemas for guiding problem solving activities. Most versions of the effect occur when less knowledgeable learners perform better after instruction with more guidance while more knowledgeable learners perform better following instruction that includes less guidance.

The main instructional implication of the expertise reversal effect within a cognitive load framework is the need to adapt instructional methods to varying levels of learner expertise (e.g., Kalyuga & Sweller, 2005; Kalyuga, 2007). Kalyuga and Sweller (2004; 2005) investigated adaptive instruction through the use of a rapid test of expertise that allowed their instructional materials to be dynamically altered as learners’ expertise changed during the experimental session. This research provided strong evidence in support of adaptive instruction with the learner-adapted instruction group significantly outperforming the randomly assigned instruction group. Rapid evaluations of learner expertise were made throughout the learning session with instructional methods changed as deemed appropriate by the underlying expertise reversal effect.

The need for learner-tailored instruction was discussed in earlier studies by Tobias (1989). The effects of different methods of adaptive learning task selection in computer-based training for air traffic control were also investigated by Camp, Paas, Rikers, and van Merriënboer (2001). In three experimental conditions, learning tasks were selected using reported mental effort, performance scores, or a mental efficiency indicator that combined these two variables while in the control group, learning tasks were provided in a fixed predetermined simple-to-complex task sequence. Results indicated that the learner-tailored task selection generated better results than non-tailored task selection, with no significant advantages of any of the task selection methods over the other two methods. Similar results were obtained by Kalyuga (2006) in a study based on the expertise reversal effect.

Isolated - interacting elements effect

The instructional material’s element interactivity is one of the key concepts in cognitive load theory (Sweller, 1994, 2010; Sweller & Chandler, 1994). Levels of element interactivity are determined by the extent to which learning elements can be assimilated individually or only in combination with other elements. When elements interact with other elements, they all must be learned simultaneously resulting in a heavy working memory load. Alternatively, if an element can be learned in isolation from other elements because it does not interact with them, then element interactivity is low. It needs to be emphasised that levels of element interactivity are not just dependent on the characteristics of the information being processed but also depend on the knowledge base of a learner. High element interactivity for a novice in an area may constitute low element interactivity for a more expert learner. Recent work by Johnson and Slayter (2012) with introductory accounting students illustrated the applicability of this concept in an accounting context. The appropriate instructional design was achieved by reducing the complexity of the learning task by initially restricting the scope of accounting transaction types.

Acquiring organised knowledge structures in long-term memory allows multiple interacting elements to be treated as a single higher-level element thus reducing cognitive load. The expertise reversal effect occurs because novice learners tend to have their working memory overloaded by highly interactive learning material that has not as yet been chunked into fewer elements. Novice learners will therefore usually benefit from well-guided instructional materials (Kirschner, Sweller, & Clark, 2006) that introduce new elements of information gradually. More knowledgeable learners, on the other hand, may have their working memory overloaded as they are required to process instructional guidance that is unnecessary for them because their partially developed schemas already provide the required guidance.

The segmenting or breaking down of a complex task to facilitate learning for novice learners is an instructional method applicable to all but the most basic physical or mental activities. However such isolated elements instruction requires an additional step in the learning process, being the subsequent integration of the learning task’s various
basic elements to achieve the ultimate learning goal. Thus, the presentation of information as isolated elements should be followed by the presentation of the same elements in an interactive form. Initially, for novice learners, isolated elements instruction may be a necessity as their working memory could be overloaded if they attempt to deal with the learning task in its entirety, with all the interacting elements involved (Pollock, Chandler, & Sweller, 2002). In contrast, for more prepared learners, the same step-by-step instruction that is beneficial to novices may impose a greater cognitive load than presenting all the interacting elements simultaneously. This increased cognitive load forms the basis of the redundancy effect according to which requiring a learner to cognitively process information that they already possess imposes an additional, extraneous cognitive load as the learner must reconcile the redundant information with what they already know. Learning could be inhibited by the need to process the redundant information when understanding is already ensured by the existing knowledge structures in long-term memory, while the learner is nevertheless provided with the same information in another form. Therefore, any forms of instructional guidance and learning activities that are essential for novices may have negative consequences for more experienced learners, especially if these learners cannot ignore or otherwise avoid such redundant information or activities.

For example, Clarke, Ayres and Sweller (2005) obtained an expertise reversal effect with instruction using spreadsheets to teach mathematics concepts. The study showed that students with low-level experience and knowledge of spreadsheets learned more effectively if they were provided with spreadsheet instruction separately before being presented with the mathematics material. However, this instructional format resulted in a redundancy effect and reduced learning efficiency for more experienced spreadsheet users.

The current study

The study reported in this paper was designed to investigate the expertise reversal effect based on the isolated-interactive elements instructional procedure in the domain of accountancy to provide more evidence for the relevance of cognitive load theory to accounting and other similar “rule-based” domains such as engineering, particularly in relation to the opportunities for its application that are currently provided by means of contemporary technology. As mentioned by Mostyn (2012), there has not been widespread application of cognitive load theory to the discipline of accounting. Halabi (2004) and Halabi et al. (2005) examined the use of worked examples in accounting education focusing on measuring the efficiency of instruction for students with differing levels of prior knowledge. They found that students who had not previously studied accounting achieved higher learning efficiency with the use of worked examples compared to problem solving exercises. Further work by Halabi (2006) with introductory accounting students showed that prior knowledge impacted on the efficiency of feedback type with rich feedback significantly more efficient for novice students.

Blayney et al. (2010) found an expertise reversal effect with accounting students that was driven primarily by the learning benefits for novices from receiving isolated elements instruction during the initial phase of learning. In accordance with the general logic of prior expertise reversal research, this paper reports the results of two studies in the domain of accounting instruction. The first experiment investigated if an expertise reversal effect could be confirmed in this domain. It extends previous research by Blayney et al. (2010) with introductory accounting students that provided evidence in support of the expertise reversal effect. Those previous results showed a significant disordinal interaction between the isolated and interactive elements learning conditions and levels of learner expertise. Isolated elements instruction eliminates explicit connections between various elements in order to reduce working memory load. Interacting elements instruction includes the connections between elements. Less prepared or lower expertise students learned more from the isolated elements instructional method, while students with greater expertise performed better with the interacting elements format.

A detailed analysis of those earlier results revealed that the effect had been driven primarily by novice learners benefiting from isolated elements instruction with no statistical differences between relatively more experienced students. A possible reason for this non-significant result for more experienced learners is that their level of expertise was not sufficiently high to render a reversal in the relative effectiveness of the isolated and interacting elements instructional methods; i.e., even the most expert learners lacked sufficient schemas to fully benefit from interacting elements instruction.
In a similar manner to this previous research, the current Experiment 1 investigated the learning effect of using intermediary solution formulas to solve complex (i.e., high element interactivity) accounting problems for learners with different levels of domain expertise. However, this experiment used students who, in contrast to participants in the Blayney et al. (2010) study, had previously been exposed to the computer laboratory environment on two occasions. Not having to learn how to function in a new learning environment may have freed working memory resources that consequently could be deployed to learn the accounting content (Clarke et al., 2005). Using these learners who generally were relatively experienced in a computer-based environment was expected to demonstrate an expertise reversal effect with a statistically significant reversal in the effectiveness of isolated and interacting instructional methods for expert learners.

The second experiment reported in this paper investigated the instructional effectiveness of tailoring instruction to levels of learner expertise based on the established expertise reversal effect. In contrast to previous studies within a cognitive load framework (e.g., Kalyuga & Sweller, 2004; 2005), Experiment 2 attempted to establish the benefits of adaptive instruction with the use of a simpler, non-dynamic assessment method. Learners’ expertise levels were evaluated once only at the beginning of the instructional session with the assigned instructional method based on levels of expertise maintained throughout the session. This approach is similar to the pre-task adaptation model suggested by Tennyson (1975) that assigned students to specific instructional treatments based on pre-task measures of prior achievement in a given domain taken before the learning session.

**Experiment 1**

According to the expertise reversal effect in cognitive load theory, an interaction was hypothesised with experts performing better using interactive over isolated element instructions and novices performing better using isolated over interactive element instructions. The participants in this study were relatively more experienced in general compared to participants in the study by Blayney et al. (2010), as indicated by a comparison of scores on the pre-session tests in both experiments. The results from the pre-session test showed participants in Experiment 1 performed at a significantly higher level (average of 2.56 out of 5, \(SD = 1.70\)) than participants in the previous study (average of 1.60 out of 5, \(SD = 1.40\)), \(F(1, 455) = 42.03, MSe = 96.87, p \leq .01, partial \eta^2 = 0.085\) (the same pre-test questions were used in both studies).

**Method**

**Participants**

Experiment 1 was conducted with 171 students enrolled in a first year university accounting course comprised of 63 (37%) males and 108 (63%) females. All students had successfully completed a prerequisite introductory accounting course and were studying to earn a Bachelor of Commerce degree (68.4%), Bachelor of Economics degree (7.6%) or a combined Commerce degree such as Bachelor of Commerce and Laws (4.1%), Bachelor of Commerce and Arts (8.2%), Bachelor of Engineering and Commerce (4.1%) and some other degrees.

Only students who felt that they needed the additional tutorial sessions were expected to attend. Participant numbers were further reduced by attending students who did not agree to participate in the study and by those who failed to attempt both the pre-test and post-test questions. As a result only 171 students (out of the total class enrolment of 429 students) actually participated in the study. Participants’ relative levels of expertise were determined from the scores achieved on five pre-test questions administered at the beginning of the tutorial session. It needs to be reiterated that in this study (as in research on the expertise reversal effect in general), expertise is considered on a continuum rather than as a dichotomy: the participants are not actual “experts” but rather relatively less or more knowledgeable learners in a specific task domain.

**Materials and procedures**

Students were presented with learning materials using a highly customized Excel computer spreadsheet designed to largely eliminate the need for specific spreadsheet skill and previous spreadsheet experience. The research
instrument used in this study incorporated the use of automated system-controlled procedures to allocate each student’s laboratory session time (i.e., 50 minutes) to the various tasks. Students’ time on task for the pre-test and post-test questions was strictly limited to 5 minutes and 10 minutes respectively. The structure of the laboratory session included the following stages:
(1) Log on to computer and open spreadsheet model - 5 min
(2) Pre-session knowledge test (five questions) - 5 min
(3) Main learning activity—phase 1 - 15 min
(4) Main learning activity—phase 2 - 15 min
(5) Post-session knowledge test (ten questions) - 10 min

The main learning activities of the lab sessions (stages 3 and 4 above) comprised 30 minutes (i.e., 60%) of the 50-minute session. Prior to this activities, students were instructed that Phase 1 would provide five problems and a bonus question (if they still had some time left), and Phase 2 repeated similar problems with different data. Students were also provided with a worked example prior to receiving each problem (to be used if they were uncertain about solving that type of problem). The participants were reminded that they had about 3 min for each of the five problems to complete the tasks in the 15 min time limit, and that for each problem, they would be given further help after three incorrect attempts. After three incorrect attempts students were provided with a dialogue box describing the solution process together with a solution formula (e.g., +D6-D8-D12).

The key feature of the automated spreadsheet was the provision of learning materials in one of two formats; isolated elements or interacting elements. Interacting elements instruction required students to provide their answers to the various accounting problems in a single spreadsheet cell. Intermediary entries (rough workings) in other spreadsheet cells were not permitted for this experimental group. Therefore, the interacting elements group participants were restricted to a single, final answer cell. In contrast, isolated elements instruction required completion of one or more “working entries” (isolated elements) before finalising a solution to the accounting problem with a formula incorporating the working entry calculations (interactive elements). For both experimental groups, comprehensive worked examples were available to assist participants in learning. Students could access these examples as desired at any time during phases 1 and 2. These worked examples could not be used during the final stage - post-question performance test.

Learning was assisted for students in both experimental groups with the provision of the three clickable buttons shown in Figure 1. In addition to the ability to display a worked example students could also self-assess their work at any time or be shown the correct numerical answer for the problem. The “Show correct answer” button did not however provide the formula required to achieve the required result. Entering of this correct answer value to the shaded entry cell yielded a “Sorry, not right yet” feedback. All entry cells required the input of a proper spreadsheet formula to be assessed as being correct.

![Figure 1. Problem displaying learning assistance features available with all Phase 1 & 2 problems](image-url)
Figure 2 illustrates an accounting problem provided to the isolated elements experimental group with its provision of (and requirement for) three intermediary or working entries. In this illustration the student has already made an attempt for each of the three working entries to cells F26, F29 and F32. Intermediary formula #1 is comprised of the entry \((E7*F6)-(D7*E6)\) which provided the value of 3 in cell F26 for the isolated elements format. Intermediary formula #2 should then have used this result plus the February sales forecast in cell \(E6\) to determine the production volume required for the month of 123 units. In the same manner intermediary formula #3 could incorporate this value multiplied by the direct material requirement of 50 (cell E12) to derive the kilograms of direct material needed for February production (i.e., 6150 in cell F32).

The complexity of the problem in Figure 2 is derived from the multiple concepts that must be understood before the arithmetical calculations are undertaken to determine the February direct material purchase requirement. Interactive elements instruction students had to recognise that firstly, they needed to calculate a production volume amount for February before any direct material values could be determined. In contrast, students in the isolated elements group were provided with a template for calculation of the production volume amount (intermediary formulas #1 & #2). Students then needed to recognise the need to convert the production volume amount into a direct material quantity. While the isolated elements group was provided with intermediary formula #3 in Figure 2, participants in the interactive elements group simply needed to know how to include this calculation in their overall formula. Both experimental groups were then required to convert the February direct material production requirement (i.e., 6150 kg) to the February purchase volume by accounting for the change in the direct material volume from the beginning to the end of the month.

With the interacting elements format, the students were not allowed to use intermediary formulas. The complete answer formula (e.g., \(=((E7*F6)-(D7*E6)+E6)*E12+E13-D13\)) had to be entered in a single spreadsheet cell, cell F23 for the Figure 2 example. Rows 25-32 containing the intermediary formulas in Figure 2 were not provided to the interacting elements instructional group. The procedure required students in this experimental group to solve the accounting problems without completing the intermediary formula working entries. All other aspects of the learning model were unchanged between the two experimental groups. Figure 3 below demonstrates the “single cell” format of the interacting elements method.
Experimental design and procedure

A 2 (instructional conditions) x 2 (levels of expertise) experimental design was implemented in this study. Prior to the lab session exactly one half of the class was randomly allocated to receive isolated elements instruction and the other half was instructed using interactive elements. The random non-participation of students in the experimental tutorial session resulted in some deviation of the actual numbers of participants in the two experimental learning conditions from exactly 50% (see bottom three rows of Table 1).

Testing for expertise equivalence by learning condition for the novice and expert subgroups was conducted. For novices there was no significant difference between learning conditions on the pre-session test scores, $F < 1$ ($M = 1.46, SD = 1.31$ for the interacting elements group and $M = 1.66, SD = 1.28$ for the isolated elements group), indicating equivalence between the two instructional groups. Similarly for the expert subgroup there was no significant difference between learning conditions on the pre-session test scores, $F < 1$ ($M = 4.40, SD = 0.50$ for the interacting elements group and $M = 4.28, SD = 0.45$ for the isolated elements group).

A set of five pre-test questions (e.g., pre-test question #1: calculate the number of units of product that need to be produced in January) provided to participants at the beginning of their laboratory session was used to gauge level of expertise in the accounting topic area to be learned. A cut-off score was subsequently used to designate students as relative novices or experts. Three or more correctly answered pre-test questions were required to classify a learner as a relative expert in this task domain.

Table 1 also displays the disparity in the learning condition / expertise cells caused by the relatively higher proportion of novices (according to the above criteria). Still, the totally random nature of student non-participation and preparation for learning should eliminate any possibility of bias in the results of the current study. The dependent variable was the post-test performance score based on student responses to a set of ten questions administered at the conclusion of the tutorial session. Correctly solved questions were assigned a score of 1 giving a maximum post-session test score of 10. Non answered questions were treated as incorrect responses, as their exclusion would upwardly bias these scores. Both the pre-test and post-test questions were developed by the senior staff of the academic teaching team for the introductory accounting course in which they were administered. Evaluation of the test questions was performed by various team members. Cronbach alpha scores of 0.79 and 0.68 respectively indicated a sufficiently high degree of internal reliability for the experiment’s post-test and an acceptable level for the pre-test.

Results and discussion

Table 1 indicates post-test means and standard deviations. A 2 (instructional condition: isolated vs. interacting elements) x 2 (levels of expertise: novice vs. expert) ANOVA indicated no overall effect of instructional condition, $F$
(1, 167) = 0.261, $MSE = 1.879$, $p = .610$, partial $\eta^2 = .002$; a significant effect of expertise, $F (1, 167) = 24.522$, $MSE = 176.413$, $p < .001$, partial $\eta^2 = .128$; and a significant interaction effect, $F (1, 167) = 5.880$, $MSE = 42.298$, $p < .05$, partial $\eta^2 = .034$. Figure 4 demonstrates the disordinal interaction between expertise and learning condition.

**Table 1.** Post-question results by expertise and instructional condition in Experiment 1

<table>
<thead>
<tr>
<th>Expertise</th>
<th>Instructional condition</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice</td>
<td>Interacting elements</td>
<td>5.15</td>
<td>2.836</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Isolated elements</td>
<td>5.98</td>
<td>2.925</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>5.64</td>
<td>2.904</td>
<td>110</td>
</tr>
<tr>
<td>Expert</td>
<td>Interacting elements</td>
<td>8.36</td>
<td>1.655</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Isolated elements</td>
<td>7.08</td>
<td>2.590</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>7.61</td>
<td>2.326</td>
<td>61</td>
</tr>
<tr>
<td>Total</td>
<td>Interacting elements</td>
<td>6.28</td>
<td>2.914</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Isolated elements</td>
<td>6.38</td>
<td>2.845</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.34</td>
<td>2.866</td>
<td>171</td>
</tr>
</tbody>
</table>

Following the significant interaction, tests for simple effects revealed a significant difference between instructional conditions favouring the interacting elements condition for the expert learners $F(1, 61) = 4.721$, $MSE = 24.047$, $p < .05$, partial $\eta^2 = .074$, but no significant difference between instructional conditions for novice learners, $F(1, 108) = 2.222$, $MSE = 18.535$, $p = .139$, partial $\eta^2 = .020$.

**Figure 4.** Interaction between levels of learner expertise and isolated-interacting learning conditions

The primary hypothesis for this experiment was that students would learn more when they receive an appropriate method of instruction according to their levels of expertise. Blayney et al. (2010) provided strong evidence for this hypothesis for novice learners but much weaker evidence for more expert learners. Experiment 1 was carried out to see if an advantage could also be obtained for relatively more knowledgeable students learning from interactive elements instruction by using generally more experienced learners than previously.

The results confirm that for experts, presenting instruction in an interacting element format was beneficial. These learners were sufficiently knowledgeable to be able to benefit from interacting elements instruction by employing their partially developed schemas. When using an isolated elements instructional format, knowledge acquisition by these learners was presumably inhibited by the inclusion of redundant learning material that did not emphasize the relations between elements that these more knowledgeable learners required.

On the other hand, in contrast to Blayney et al. (2010), for novices, while the means were in the expected direction with the isolated elements condition resulting in higher test scores than the interacting elements condition, the difference was not statistically significant. Since the participants in Experiment 1 were generally more experienced than learners in the previous study, the levels of expertise of the novice participants may have been higher than that
required to demonstrate a benefit of using isolated over interacting elements format. Together, the results of this experiment and those of Blayney et al. (2010) imply the need to tailor instructional formats to levels of learner expertise in this specific instructional domain.

**Experiment 2**

In accordance with the expertise reversal effect and the results of Experiment 1, less knowledgeable students should learn better when instruction is presented in an isolated elements form while more knowledgeable students should learn better when instruction is presented in an interactive elements form. The main hypothesis of Experiment 2 is that for a group of students that includes learners at different levels of expertise, learning will be facilitated if students are provided with instruction that is tailored to their level of expertise (i.e., adaptive instruction). Experiment 2 used the above adaptation principle based on the expertise reversal effect for the allocation of participants to the two alternative instructional formats in the study’s adaptive condition group. The allocation of the non-adaptive group’s participants between instructional formats was done on a random basis irrespective of their levels of expertise.

The random allocation of instructional format (i.e., isolated or interacting elements) to participants in the non-adaptive group was expected to cause relatively lower learning outcomes for participants who did not receive the optimal instructional format according to the expertise reversal effect. In contrast to the non-adaptive group, participants in the adaptive group received the optimal instructional format according to their level of expertise. Accordingly, the overall expectation was that students in the adaptive group would perform better on the post-test questions at the end of their laboratory session than their counterparts in the non-adaptive group. Thus, this experiment tested the effectiveness of the adaptation methodology based on the previously established expertise reversal effect in this specific domain.

**Method**

**Participants**

Experiment 2 was conducted with 94 students enrolled in a first year university accounting course comprised of 39 (41%) males and 55 (59%) females. As in Experiment 1, all students had successfully completed a prerequisite introductory accounting course and were studying to earn a Bachelor of Commerce degree (76.6%), Bachelor of Economics degree (3.2%) or a combined Commerce degree such as Bachelor of Commerce and Laws (8.5%), Bachelor of Commerce and Arts (3.2%), Bachelor of Engineering and Commerce (2.1%) and some other degrees. Only students who felt that they needed to attend the additional tutorial sessions were expected to participate. As a result, in conjunction with non-agreement to participate in the study and failure to attempt both the pre-test and post-test questions, only these 94 students (out of the total class enrolment of 299 students) actually participated in the study.

The attending student group was further reduced by nine (to 85) by eliminating those participants who took less than 120 seconds to complete the five pre-questions administered at the beginning of their tutorial laboratory session. Given that it took a domain expert more than two minutes to complete the five pre-questions it was considered that these participants had not given an honest effort to this task thus making it impossible to judge their level of expertise. Since the level of prior knowledge was used as the basis for allocation of participants to learning condition sub-groups, the data for these nine participants was removed from the analysis.

**Materials and procedures**

Students were presented with learning materials using a customised, largely automated Excel computer spreadsheet similar to that used in Experiment 1. The automated spreadsheet provided the learning materials in one of two formats: isolated elements (multiple spreadsheet cells) or interacting elements (single spreadsheet cell). Students were assigned to one of these learning formats based on their levels of prior knowledge in the adaptive condition or
randomly in the non-adaptive condition. The same automated system-controlled procedures as those used in Experiment 1 determined and equalised students' time on task in Experiment 2.

**Experimental design and procedure**

The 85 participants were randomly assigned to either an experimental adaptive instruction group (46 participants) in which less knowledgeable learners were presented isolated elements instruction and more knowledgeable learners were presented interactive elements instructions or a non-adaptive control group (39 participants) in which learners were allocated to isolated or interactive instructional formats on a random basis. An unequal split between the groups resulted from students' random non-attendance to the tutorial sessions. There was no significant difference between the adaptive and non-adaptive groups on the pre-session test scores, $F < 1$ ($M = 1.91, SD = 1.41$ for the adaptive group and $M = 1.79, SD = 1.58$ for the non-adaptive group), indicating equivalence between the two instructional groups.

Participants in this study completed pre-session and post-session knowledge tests identical to those used in Experiment 1. The learning condition allocation scheme was programmed into the computer spreadsheet tutorial with the novice/expert cut-off designation based on the participant's performance on the pre-test questions administered at the beginning of their laboratory session. Three or more correct pre-test questions were required for a participant to be deemed a relative expert (more knowledgeable student). Participants achieving less than three correct pre-test questions were deemed novices.

**Results and discussion**

The dependent variable was the post-test performance score based on student responses to a set of ten questions administered at the conclusion of the tutorial session, while the independent variable was the instructional condition (adaptive vs. non-adaptive). An ANOVA for the two instructional groups revealed higher performance for the adaptive group ($M = 7.17, SD = 2.48$) compared to the non-adaptive group ($M = 5.97, SD = 2.71$), $F(1, 84) = 4.54$, $MSE = 30.37$, $p < 0.05$, partial $\eta^2 = .020$.

This study's main hypothesis was that students in the adaptive instruction condition who received the appropriate method of instruction according to their level of expertise would perform at a higher level than students in the non-adaptive instruction condition who were randomly assigned to an instructional method. The results supported this expectation by indicating that adaptive instruction that took the learner's expertise into account was beneficial.

This study demonstrated that learning could be improved through the adaptation of instructional techniques according to the expertise of the individual even with a very simple adaptation approach based on using only pre-test results for assigning students to instructional techniques. Less knowledgeable learners need to be compensated for their lack of relevant schemas in long-term memory through the provision of an isolated elements learning environment, while more knowledgeable learners may be able to successfully deal with fully interactive instructional materials.

**General discussion**

The major finding of Experiment 1 concerns the interaction between the isolated-interacting elements effect and levels of learner expertise. Previous studies have indicated that for novices, presenting very high element interactivity information in isolated rather than in its more natural, interactive form during the initial phase of instruction is beneficial. According to cognitive load theory, novices may have difficulty processing very complex information in the initial stages of instruction and so should benefit from having the information initially divided into isolated elements. That result was obtained in previous studies, including Blayney et al. (2010) in accounting instruction and Pollock et al. (2002) in engineering instruction. In those studies, it also was hypothesised that providing more knowledgeable learners with interactive rather than isolated information should be beneficial. More knowledgeable learners may have already acquired the knowledge associated with individual elements. Providing that knowledge during instruction may be redundant with all of the negative consequences associated with the
redundancy effect. Rather than learning about individual elements of information, more expert learners may need to learn how various elements of information interact. Presenting information in interactive form is more likely to demonstrate to students how elements interact than if those elements are presented in isolated form. Previously, that result was not obtained but the current study fills that particular gap. Experiment 1 demonstrated that for more knowledgeable learners, presenting information in interactive form was superior to presenting it in isolated form, a result that was not obtained for novices.

The distinction between the results of Blayney et al. (2010) and the results of Experiment 1 of the current study can be ascribed to different levels of expertise between the two relevant populations. The generally relatively more knowledgeable learners in Experiment 1 of the current study might have had sufficient working memory capacity to easily process isolated elements and rather, required information on the manner in which the elements interact. That information could be obtained more readily by these learners from the interactive elements condition than the isolated elements condition. In contrast, the less knowledgeable learners of Blayney et al. (2010) required information presented in isolated elements form. Overall, these results are in line with other studies on the expertise reversal effect (Kalyuga, 2007; Sweller et al., 2011).

An implication of this result is that better, more finely-grained methods of evaluation of learner prior experience are required for optimal tailoring of instructional methods to levels of learner expertise. However, even a simple, static pre-session allocation of learners to corresponding instructional methods (the pre-task adaptation model according to Tennyson, 1975) in Experiment 2 was sufficient to demonstrate the significant benefits of learner-tailored instruction. From a practical, instructional perspective, the current results in conjunction with previous findings indicate the importance of the isolated-interacting elements effect. When providing novices with information that is very complex, it is likely to be desirable to decompose that information into isolated elements in the first instance. While isolated elements may only result in a partial understanding of the information, that negative consequence can be rectified by following the partially understood information with fully interactive materials allowing full understanding. Nevertheless, as the current work demonstrates, if isolated elements instruction is provided to learners who have already assimilated those elements, the consequences will be negative rather than positive. More knowledgeable learners should be presented with fully interactive information at the beginning of instruction. Based on these conclusions, the need to use instruction that is adaptive to levels of expertise is critical.

It should be noted that, even though the simple pre-task form of adaptation used in Experiment 2 improved learning in comparison with non-adapted instruction, this form of adaptation may not work with relatively more complex or multi-task tutoring sessions that require multiple, real-time adaptations based on the assessment of learner progress during the tutorial. Various specific techniques have been proposed and developed to implement the concept of adaptive instruction. Renkl and Atkinson (2003) suggested a faded worked example approach for transitioning from fully worked-out examples to conventional problems. Kalyuga and Sweller (2004; 2005) proposed the use of rapid diagnostic tests for monitoring learner’s progress and altering the instructional techniques in real time. Such adaptive learning environments that dynamically tailor instructional formats to changing levels of learner expertise need to be investigated in future studies using the potential of contemporary technological means. Future research may also be enhanced through the establishment of internal validity by measuring and analysing indicators of participants’ cognitive load during the learning and testing phases and possibly using such indicators for refining adaptation procedures.

References


The Effects of Social Cue Principles on Cognitive Load, Situational Interest, Motivation, and Achievement in Pedagogical Agent Multimedia Learning

Sanghoon Park
Department of Educational and Psychological Studies, University of South Florida, USA // park2@usf.edu

ABSTRACT
Animated pedagogical agents have become popular in multimedia learning with combined delivery of verbal and non-verbal forms of information. In order to reduce unnecessary cognitive load caused by such multiple forms of information and also to foster generative cognitive processing, multimedia design principles with social cues are suggested (Mayer, 2014a). This study presents the design model of pedagogical agent multimedia learning by using four design principles based on social cues. Then reported are the findings of a study examining the effects of the pedagogical agent multimedia learning on perceived cognitive load, situational interest, motivation, and achievement. One hundred twenty seven college undergraduate students enrolled in four “Computer literacy” classes participated in this study and were randomly assigned to one of the six conditions in which on-screen images (image vs. no image) and narrations of a pedagogical agent (human voice narration, on-screen text, no narration) were presented in different levels. The results indicated that, overall, the presence of images does not significantly affect perceived cognitive load, situational interest, motivation, or achievement. However, the form of narration influenced the four outcome measures differently. The use of human voice narrations presented by a pedagogical agent was effective to reduce the perceived cognitive load compared to on-screen text narration and no narration conditions. Human voice narration by a pedagogical agent was also found to promote learners’ situational interest, which is negatively correlated to cognitive load. The personalized narration was found to improve learners’ motivation in terms of relevance and confidence whether presented by a pedagogical agent or in on-screen text although no significant differences were found in the recall test and the comprehension test.

Keywords
Cognitive load, Social cues, Generative processing, Pedagogical agent, Multimedia learning, Motivation

Introduction

The use of animated pedagogical agents in multimedia learning environments has increased as new technologies have made them more accessible (Gholson & Craig, 2002; Johnson, Rickel, & Lester, 2000). Pedagogical agents are animated life-like characters enabled with speech, gesture, movement, and human-like behaviors (Sweller, Ayres, & Kalyuga, 2011) and designed to facilitate learning in multimedia learning environments (Johnson et al., 2000). Pedagogical agents can embody different pedagogical roles to support learners by supplanting, scaffolding, coaching, testing, or demonstrating or modeling a procedure (Schroeder & Adesope, 2014). Previous studies have claimed the positive influences of pedagogical agents on student motivation and interest (Atkinson, 2002; Moreno, 2005) and also have indicated positive effects on student attitude toward learning and performance (Baylor, 2002a, 2002b; Baylor & Ryu, 2003; Moreno et al., 2001). However, other studies reported that pedagogical agents in multimedia learning could cause unnecessary cognitive load (Choi & Clark, 2006; Clark & Choi, 2005) called extraneous cognitive load. For example, the split-attention effect could occur when multiple sources of information are presented in split-attention without being integrated (Ayres & Sweller, 2014). The modality effect also can be caused when multiple sources of information are presented in single-modality not in dual modality (Low & Sweller, 2014). Both split-attention effect and modality effect are considered cognitive load effects caused by information presented through multiple information sources.

According to cognitive load theory (CLT), our brain utilizes two primary types of memory, the working memory and the long term memory, to process, store, and access information (Kalyuga, 2011; Sweller, 2005; Sweller, van Merriënboer, & Paas, 1998). Due to the limited capacity of the working memory, learners must cope with a certain level of cognitive load to process newly presented information. In pedagogical agent multimedia learning, both auditory and visual channels of information can be engaged in working memory from two different sources that are pedagogical agent and on-screen multimedia material, thus influence learners’ cognitive load as a whole. Cognitive load theorists (Pass, Renkl, & Sweller, 2003; Sweller, 1999, 2005) agree that three different types of cognitive load need to be considered in designing instruction: (a) intrinsic cognitive load, (b) extraneous cognitive load, and (c) germane cognitive load. Intrinsic cognitive load is imposed by the intrinsic nature of presented information or learning task itself and should be reduced (i.e., task difficulty). Extrinsic cognitive load results from the ineffective
instructional design and needs to be prevented (e.g., format of instructional materials). Germaine cognitive load is also imposed by instructional design but is effective for learning (e.g., a learner’s effortful process of understanding). The distinction between intrinsic cognitive load and germaine cognitive load is not clearly made (see Sweller, Ayres, & Kalyuga, 2011) because germaine cognitive load is not imposed by the nature and structure of the learning materials. However, germaine cognitive load has been associated with various additional cognitive activities that are designed to foster schema acquisition (Kalyuga, 2010). Therefore, it would be reasonable to consider germaine cognitive load as sources of auxiliary cognitive activities to enhance learning outcomes or to increase learner motivation (Kalyuga, 2010).

Applying the types of cognitive load to designing multimedia learning environments, Mayer (2009, 2014b) suggested three kinds of cognitive processing demands in his cognitive theory of multimedia learning (CTML) that are extraneous processing, essential processing, and generative processing (also see Moreno & Mayer, 2010). Each of the kinds corresponds to each type of aforementioned cognitive loads (Mayer, 2014b). According to the CTML, people learn with multimedia presentations based on three assumptions. First, people learn by processing visual/pictorial material and auditory/verbal material through separate channels. Second, people learn by processing limited amount of information in each channel at one time. And lastly, people learn by actively processing cognitive resources during learning, including selecting relevant information, organizing selected material into a coherent mental representation, and integrating incoming material with existing knowledge (Mayer, 2014b). Based on a central tenet common to CLT and the CTML, Mayer (2014b) summarized three demands on cognitive capacity during multimedia learning (Table 1). The Table 1 presents three types of demands on learners’ information processing during learning using the terminology of the CTML and CLT.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Caused by</th>
<th>Learning processes</th>
<th>Example</th>
<th>Cognitive load</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraneous</td>
<td>Cognitive processing that is not related to the instructional goal</td>
<td>Poor instructional design</td>
<td>None</td>
<td>Focusing on irrelevant pictures</td>
<td>Analogous to extraneous cognitive load</td>
</tr>
<tr>
<td>processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential</td>
<td>Cognitive processing to represent the essential presented material in working memory</td>
<td>Complexity of the material</td>
<td>Selecting</td>
<td>Memorizing the description of essential processing</td>
<td>Analogous to intrinsic cognitive load</td>
</tr>
<tr>
<td>processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generative</td>
<td>Cognitive processing aimed at making sense of the material</td>
<td>Motivation to learn</td>
<td>Organizing and integrating</td>
<td>Explaining generative processing in ones’ own words.</td>
<td>Analogous to germaine cognitive load</td>
</tr>
<tr>
<td>processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Strategies and techniques to meet each of the demands have been empirically tested in a number of studies. As a result, Mayer (2014b) suggested three categories of multimedia design principles based on three different learning scenarios. For example, the coherence principle, the signaling principle, the redundancy principle, the spatial contiguity principle, and the temporal continuity principle were suggested as principles for reducing extraneous cognitive load. For managing intrinsic cognitive load, suggested were the segmenting principle, the pretraining principle, and the modality principle. The third category of principles for fostering germaine cognitive load includes the multimedia principle, the personalization principle, the voice principle, the embodiment principle, the guided discovery principle, the self-explanation principle, and the drawing principle (for details about each principle, see Mayer, 2014b).

In pedagogical agent multimedia learning, many of the aforementioned principles have been studied because of the nature of pedagogical agent and the instructional messages presented through verbal and non-verbal forms. For example, the signaling principle was tested as a strategy to reduce extraneous cognitive load by manipulating a pedagogical agent’s movement and pointing gesture (Choi & Clark, 2006; Johnson, Ozogul, Moreno, & Reisslein, 2013). Also the modality principle in pedagogical agent multimedia learning has been examined was a way to manage intrinsic cognitive load by presenting information through two different sources, voice narration or on-screen text. (Moreno, 2002; Moreno, Mayer, & Lester, 2000; Moreno, Mayer, Spires, & Lester, 2001). The study findings were congruent with previous findings on modality effects in multimedia learning (Mayer, 2009; Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995). Germaine cognitive load in CLT represented as generative cognitive
processing in Mayer’s CTML, has been studied by manipulating germane sources of load (Renkl, Atkinson, & Große, 2004; Gerjets & Hesse, 2004; Berthold & Renkl, 2009). Several studies examined the learning effects of combining strategies of reducing extraneous load and increasing germane load so that students’ cognitive resources can be redirected from irrelevant to relevant schema acquisition activities (Seufert & Brünken, 2006; Seufert, Jänen, & Brünken, 2007). However, few studies have been conducted on the strategies for generative cognitive processing aimed to foster germane cognitive load in pedagogical agent multimedia learning. Recently, Mayer and Estrella (2014) conducted a study to determine the effects of emotional design features in a multimedia lesson, but their study did not involve pedagogical agent multimedia learning.

Pedagogical agent multimedia learning model with social cue-based multimedia design principles

The goal of fostering generative cognitive processing is to increase germane cognitive load and engage students into learning by organizing and integrating information presented by a pedagogical agent and on-screen multimedia material. Based upon the framework of social cues in multimedia learning (Figure 1), and the four design principles derived from the framework, the researcher developed the pedagogical agent multimedia learning model.

**Multimedia principle**

The multimedia principle supports the notion that learning with words and pictures is more effective than learning with words alone (Butcher, 2014). However, the principle currently refers more broad forms of visual, verbal, and textual content and provides a context for research examining the optimal design of multimedia learning materials (Butcher, 2014). Hence, it becomes critical to examine the conditions when and how the multimedia principle best applies (Butcher, 2014). In other words, the multimedia principle must be considered in terms of the content and the roles of images and text, and how they support or interact each other.

**Personalization principle**

The personalization principle explains that people learn more deeply when the messages in multimedia learning are designed in conversational style than formal style (Mayer, 2014b). Conversational style can be designed by using words such as “you” and “I” or by presenting the instructor’s direct-self revealing comments to learners (Moreno & Mayer, 2000; Mayer, 2014b). Previous studies used personalized messages in the form of on-screen text or voice narration (Moreno & Mayer, 2000, 2004).

**Voice principle**

The voice principle is that people learn more deeply when the words in a multimedia message are delivered in a human voice than in a machine voice (Mayer, 2014b). It supports a sense of social presence, hence human voice helps a learner feel a social response to the presented message.
Embodiment principle

The embodiment principle is that people learn more deeply when a pedagogical agent on-screen presents humanlike gesturing, movement, eye contact, and facial expressions (Mayer, 2014b). The on-screen agent could be designed in the form of a static image that shows no movement or could be designed to exhibit facial expressions, gestures, and movements.

Figure 2 illustrates the model for pedagogical agent multimedia learning with two sources of information influencing learners’ generative cognitive processing and learning. According to social cue principles, first, a pedagogical agent’s personalized human voice narration and human-like gestures/expressions stimulate learners’ interest and motivation. Second, multimedia materials (images and texts) presented on-screen also affect the learners’ interest and motivation. The instigated learning interest and motivation further promote the use of generative cognitive processing, which is analogous to germane cognitive load in CLT.

![Figure 2. Pedagogical agent multimedia learning design model based on social cue principles](image)

As depicted in the model, a rationale for implementing social cue principles in pedagogical agent multimedia learning is that both pedagogical agent and on-screen materials are intended to increase learner’s interest and motivation. Therefore, the concepts of interest and motivation need to be discussed in detail.

Interest and motivation

Individual interest and situational interest

According to Krapp (2002), interest is a relational construct that consists of a more or less enduring relationship between a person and an object. This relationship is recognized by specific activities that may comprise concrete actions and abstract mental operations. Consequently, the concept of interest may range from a single, situation-specific person-object relation (conceptualized as situational interest) towards the development of value beliefs in particular domains (conceptualized as individual interest). From the view of individual interest, interest is implied as a characteristic of person. It is specific to individuals, developed slowly, tends to be long lasting, and is triggered by an individual’s predisposition (Renninger et al., 1992; Schiefele, 1998; Schraw & Lehman, 2001; Silvia, 2001). For example, learners who are already interested in a topic or an activity pay more attention and acquire more knowledge than participants without such interest. Schiefele (1991, 1999) explained this with two subcomponents of individual interest; a feeling-related valences and a value-related. Although individual interest can be assessed through a learner analysis process by asking several background questions or administering a simple questionnaire (Keller, 1983), it is not easy to incorporate strategies to improve individual interest in designing learning material because individual interest refers to a student’s relatively enduring preference for different topics, tasks, or contexts that has been developed for a substantial period of time (Hidi & Renninger, 2006; Krapp, 1999; Tobias, 1994).
Situational interest is a prior foundation of individual interest (Krapp, 1999; 2002). When contents of a learning material is not a subject area in which the learner has established individual interest, the interesting factors in the subject learning situation is necessary to awake the interest for a short or longer period of time. The central psychological process “Internalization” supports the transformation process of situational interest into long-lasting individual interest as described in Figure 3. Situational interest is generated as a result of interestingness of situation. It is caused primarily by certain conditions and concrete objects in the environment, triggered by environmental factors, elicited by certain aspects of a situation, and it is assumed to contribute to the interestingness of the situation (Harp & Mayer, 1997; Hidi & Anderson, 1992; Krapp, 1999, 2002; Renninger et al., 1992; Schraw & Lehman, 2001; Silvia, 2001). While individual interest is a relatively stable evaluative orientation towards certain domains, situational interest is formed if an emotional state aroused by specific features of an activity or a task. Figure 3 illustrates the relationship between individual interest and situational interest.

Building upon previously conducted research on interest and development, Hidi and Renninger (2006) proposed four-phase model of interest development that describes phases of situational interest and individual interest in terms of affective and cognitive processes. The four-phase model provides a rationale for identifying early phases of interest development in terms of affect or liking. The model offers description of each phase, information about the type of support that a person in each phase of interest typically needs, and possible ways to design educational or instructional conditions to support interest development from situational interest to individual interest. The first two phases explains situational interest (triggering situational interest and maintaining situational interest). Then the last two phases suggest individual interest (emerging individual interest and well-developed individual interest).

**ARCS Motivation model**

In multimedia learning, motivation refers to how a learner initiates, energizes, maintains goal directed behaviors, and exerts effort to make sense of the instructional messages (Mayer & Estrella, 2014). Keller (2010) defined it with two elements, “direction” and “magnitude” of behaviors, and explained that motivation helps people choose goals to pursue and actively and intensively engaged them while pursuing the goals. According to Keller (2010), the concept of motivation is sorted into four categories that are Attention, Relevance, Confidence, and Satisfaction. Table 2 shows the definition of each category of the ARCS model. Among the four categories, attention refers to stimulating and sustaining a learner’s curiosity and interest while studying a lesson. People are more interested in specifics rather than in abstractions, hence using specific people or illustrating ideas with concrete examples or visualizations such as stories or images can be used to gain a learner’s attention.

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>Capturing the interest of learners; stimulating the curiosity to learn</td>
</tr>
<tr>
<td>Relevance</td>
<td>Meeting the personal needs/goals of the learner to effect a positive attitude</td>
</tr>
<tr>
<td>Confidence</td>
<td>Helping the learners believe/feel that they will succeed and control their success</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Reinforcing accomplishment with internal and external rewards</td>
</tr>
</tbody>
</table>

Gaining situational interest and attention is the first step for students to be motivated. Keller (2010) further argued that achieving the other three categories, relevance, confidence, and satisfaction, is essential for learning motivation in a lesson. Because of the commonalities between situational interest and attention, this study considers attention as an outcome of situated interest derived from interacting with text and images presented on-screen.
Research purpose and questions

As Mayer (2014a) pointed out, generative cognitive processing is essential in learning because intended learning outcomes may not be achieved if a learner experiences “generative underutilization.” Generative underutilization occurs when a learner has cognitive capacity available for generative processing but does not exert the effort to engage in learning due to the insufficient interest or lack of motivation to engage in germane cognitive activities (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Gerjets, Scheiter, & Catrambone, 2004). Mayer (2014a) suggested two types of design techniques to foster generative processing and reduce the possible “generative underutilization” problem: instructional design techniques and learning strategies. Instructional design techniques guide multimedia learning designing principles based on social cues (Mayer, 2014a) and they include the multimedia principle, the personalization principle, the voice principle, and the embodiment principle. On the other hand, learning strategies focus on supporting learners’ cognitive involvement by implementing cognitive learning activities based on the guided discovery principle, the self-explanation principle, and the drawing principle. According to Moreno (2006), however, adding learning strategies to cognitive activities has not been found necessarily helpful for learners. Since then, cognitive activities have been studied widely in multimedia learning (Mayer, 2009; Sweller, Ayres, & Kalyuga, 2011) yet, social considerations that affect the learners’ interest and motivation to engage in cognitive processing has not been emphasized much despite its’ critical role in learning (Krämer, 2010; Mayer, Fennel, Farmer, & Campbell, 2004).

In order to ensure that a learner is actively involved in generative cognitive processing and fosters germane cognitive load, it is essential that the learner is ready and willing to spend his cognitive resources on generative processing (Moreno & Mayer, 2010) so that he can produce better achievement (Mayer & Estrella, 2014). The present study examined the effectiveness of the four social cue-based multimedia design principles - the multimedia principle, the personalization principle, the voice principle, and the embodiment principle - integrated in pedagogical agent multimedia learning on learners’ both cognitive and affective aspects of learning. Specifically, this study attempted to measure interest and motivation separately based on the interest theory and the ARCS motivation theory. According to Hidi and Renninger (2006) and Keller (2010), a learner’s “willingness” to exert an effect begins with “interest.” Hidi and Renninger (2006) suggested in their interest development model that “situational interest” has to be triggered and maintained before being developed to “individual interest.” Keller (2010) in his ARCS motivation model also explained that “capturing interest,” “stimulating inquiry,” and “maintaining attention” are the first steps that influence motivation to learn. Therefore, measuring situational interest separately from motivation will help understand how two different development stages of motivation affect cognitive load differently.

Based on the pedagogical agent multimedia learning design model depicted in Figure 2, the researcher developed a Web-based learning material on Intellectual property. Then the effects of social-cue principles applied to the design of learning material were examined on learners’ perceived cognitive load, situational interest, motivation, and achievement. The following research questions were formulated to explore this issue.

RQ1. What is the effect of multimedia design principles using social cues on cognitive load in pedagogical agent multimedia learning?
RQ2. What is the effect of multimedia design principles using social cues on situational interest in pedagogical agent multimedia learning?
RQ3. What is the effect of multimedia design principles using social cues on motivation in pedagogical agent multimedia learning?
RQ4. What is the effect of multimedia design principles using social cues on achievement in pedagogical agent multimedia learning?

Method

Research design

The study used a 2 × 3 factorial design as shown in Table 3. The study variables included the use of images (presence vs. absence) to test the multimedia principle and the source of narrations (Human voice narration delivered by a pedagogical agent, text narration delivered on-screen, no narration) to test the social cue principles including the personalization principle, the voice principle, and the embodiment principle. This study employed a randomized
group post-test design. In order to explore the established research questions, participants were randomly assigned into the one of the six conditions (A through F) based on the sequence of their entry to the research lab.

<table>
<thead>
<tr>
<th>Group</th>
<th>Random assignment</th>
<th>Condition</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer literacy class students</td>
<td>R</td>
<td>A</td>
<td>X1 Y1</td>
<td>O1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>X1 Y2</td>
<td>O2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>X1 Y3</td>
<td>O1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>X2 Y1</td>
<td>O4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>X2 Y2</td>
<td>O5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>X2 Y3</td>
<td>O6</td>
</tr>
</tbody>
</table>

*Note.* a First independent variable: the use of graphic (1: presence, 2: absence); b Second independent variable: the source of narration; (1: pedagogical agent, 2: text narration, 3: no narration).

**Participants**

The participants were 127 college undergraduate students enrolled in “Computer literacy” classes in a large public university located in the southeastern United States. This course was one of the required courses for the undergraduate students. All participants were recruited from ten sections of the course and offered extra credits as compensation. Only participants who agreed to participate in the study by signing the consent form were included in the final data analyses. A total of 136 students voluntarily participated at the beginning of the study. Of these, 127 participants were included in the final data analyses because nine students didn’t complete the post questionnaire. All of the 127 participants were undergraduate students. The average age of the sample was 19.72 years (SD = 1.96). Among those 127 participants, 63.0% were Caucasian, 19.7% were African-American, 11.0% were Hispanic/Latino, 2.4% were Asian/Asian American, 1.6% were bi-racial, and 2.3% were other ethnicity groups. There were 60.6% of male students and 39.4% of female students. The majority of the participants were sophomores (43.3%) with 21.3% freshmen, 20.5% juniors, and 15.0% seniors.

**Research material**

The topic of instructional material used in this study was “Introduction to intellectual property.” The topic consisted of three sub concepts of intellectual property including patent, trademark, and copyright. “Intellectual property” was selected for several practical reasons and theoretical reasons. First, the topic “Intellectual property” had been one of the topics in the Computer literacy class. Therefore, students were aware of the topic, but not familiar with detailed information, specifically in the areas of patents, trademarks, and copyrights. Second, this topic was related to the students’ everyday life regarding how to use computer applications without violating any legal and ethical issues. Third, in order to fulfill the primary goal of this study and investigate the effect of social cue principles in pedagogical agent multimedia learning, instructional text for which students presumably had low levels of prior knowledge and low interest was used.

Instructional material on Intellectual property consisted of three learning phases: (1) Introduction phase: Students were given a brief introduction about intellectual property and basic information containing history, related regulations, and examples. (2) Learning phase: Students were given a detailed explanation in regard to the three sub-concepts of intellectual property: patent, trademark, and copyright. (3) Test phase: After completing instructional material, students were given an opportunity to actually test what they had read in the instructional material.

The researcher developed a multimedia lesson consisting of 14 screen pages for each condition with two introduction slides (s1-s2), 11 learning material slides (s3-s13) and one test slide (s14). Each of the 11 learning material slides consisted of two types of information: expository information and supplementary information because the focus of the lesson presented in this study was concept learning. Three different types of “intellectual property” were presented on-screen in expository texts (Figure 4). Then supplementary texts were designed based on the personalization principle to provide examples and cases of copyright, trademark, and patent in a real life (Figure 5) and delivered in a human voice narration or on-screen text depending on the study condition. Unlike scientific expository text, learners in concept learning can form prototypes and exemplars of the concept by encountering
varied instances of the concept (Ormrod, 2012). By presenting definitions and examples hand in hand, more effective concept learning can occur (Dunn, 1983; Tennyson, Youngers, & Suebsonti, 1983).

A patent as one of three main types of intellectual property gives an inventor the right for a limited period to stop others from making, using or selling an invention without the permission of the inventor. It is a deal between an inventor and the state in which the inventor is allowed a short term monopoly in return for allowing the invention to be made public. There are some special features of a patent in terms of its scopes. In addition, specific conditions must be fulfilled to get a patent. Patents as one of three main types of intellectual property are about functional and technical aspects of products and processes. Most patents are for incremental improvements in known technology - evolution rather than revolution. The technology does not have to be complex. Patent rights are territorial; a US patent does not give rights outside of the US. Patent rights last for up to 20 years in the US.

Here’s something interesting… Do you drive your own car? You’ve probably never wondered who invented “power steering valves”? Well, let me tell you. Bishop power steering valves are used in about one fifth, or 20%, of the world’s cars. A.E. Bishop & Associates, the company that makes the valves, receives a royalty payment of up to $1.00 for each unit someone manufactures and generates its financial return through effective use of intellectual property. The story goes that Arthur Bishop began developing and patenting automotive power steering systems after World War II. His company now has over 300 patents worldwide AND it makes about $5 million a year in royalties. This means that his company doesn’t manufacture any products itself but licenses car and component manufactures and specialist manufactures of machinery around the world to use its designs. Obviously they make sure their patents are vigorously defended. And I can tell you that they prosecute many infringers. Infringers are people who try to use the valve without paying the royalty.

The total number of images used in the material was 10, one in each of the slides s4-s13, which was presented in the identical way across the conditions A-C. The total word count of expository on-screen text was 1021 with an average of 92.82 in each of the slides s3-s13. The content of the expository on-screen text was identical across the conditions A-F. The total word count of supplementary text delivering the personalized narration (whether voice or on-text) was 1377 with an average of 125.18 in each of the slides s3-s13. The content of the personalized narration was identical in conditions A, B and D E. The summary of multimedia elements and design principles used in the six study conditions are presented in Table 4.

Table 4. Multimedia elements and design principle for study conditions

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Multimedia elements</th>
<th>Total number of elements</th>
<th>Average time (in minutes) taken to finish lesson</th>
<th>Design principles applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>Cartoon image</td>
<td>10 images 1021 words</td>
<td>10.38 min</td>
<td>Multimedia principle</td>
</tr>
<tr>
<td></td>
<td>Expository information (on-screen text)</td>
<td>1377 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Supplementary information (Human voice narration)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition A</td>
<td>Voice narration with pedagogical agent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition B</td>
<td>On-screen text</td>
<td>10 images 1021 words</td>
<td>11.63 min</td>
<td>Personalization principle / Embodiment principle / Voice principle (Pedagogical agent)</td>
</tr>
<tr>
<td>Condition C</td>
<td>No narration or text</td>
<td>10 images 1021 words</td>
<td>11.76 min</td>
<td>Multimedia principle</td>
</tr>
</tbody>
</table>

218
In order to maximize the positive effect of combining auditory and visual information and also to minimize the influences of extraneous cognitive load caused by poor instructional design (Mayer, 2014b) such as a reverse modality effect, the instructional material was presented in a self-paced format. It also ensured that learners have time to transfer information from working to long-term memory without being affected by a working memory overload (Sweller, Ayres, & Kalyuga, 2011). Information on each screen was presented in the order of “expository information” to “supplementary information” to ensure that learners read or listen to the information in the same sequence. When supplementary information was presented in the form of human voice narration, a “Listen” button was shown on the Web material for students to click on to listen to the human voice narration after reading the expository information. Hence, students followed the sequence of “reading expository information” to “listening to agent delivering supplementary information.” When supplementary information was presented in the form of on-screen text, students were presented with the expository information first, then the supplementary information was presented afterward in the form of on-screen text. Therefore, students followed the sequence of “reading expository information” to “reading supplementary information.” Students were directed to go through all screens in the order from the introduction phase to the test phase using the “Next” button to prevent students from referring to the previous text while taking a test.

Independent variables

All six conditions used the same expository text on intellectual property. Then two independent variables, (1) images supporting the expository text and (2) the sources of narration, were manipulated differently in each of the six conditions. Images were intended to trigger learners’ interest in the expository text and also to help learners acquire the meaning of the concept explained in each screen. For this purpose, one cartoon image was presented on each screen in two levels (presence vs. absence).

In order to maximize the positive effect of combining auditory and visual information and also to minimize the influences of extraneous cognitive load caused by poor instructional design (Mayer, 2014b) such as a reverse modality effect, the instructional material was presented in a self-paced format. It also ensured that learners have time to transfer information from working to long-term memory without being affected by a working memory overload (Sweller, Ayres, & Kalyuga, 2011). Information on each screen was presented in the order of “expository information” to “supplementary information” to ensure that learners read or listen to the information in the same sequence. When supplementary information was presented in the form of human voice narration, a “Listen” button was shown on the Web material for students to click on to listen to the human voice narration after reading the expository information. Hence, students followed the sequence of “reading expository information” to “listening to agent delivering supplementary information.” When supplementary information was presented in the form of on-screen text, students were presented with the expository information first, then the supplementary information was presented afterward in the form of on-screen text. Therefore, students followed the sequence of “reading expository information” to “reading supplementary information.” Students were directed to go through all screens in the order from the introduction phase to the test phase using the “Next” button to prevent students from referring to the previous text while taking a test.

Independent variables

All six conditions used the same expository text on intellectual property. Then two independent variables, (1) images supporting the expository text and (2) the sources of narration, were manipulated differently in each of the six conditions. Images were intended to trigger learners’ interest in the expository text and also to help learners acquire the meaning of the concept explained in each screen. For this purpose, one cartoon image was presented on each screen in two levels (presence vs. absence).
Narrations were presented in three different levels. First, personalized human voice narration was delivered by a pedagogical agent. Second, personalized text was presented on-screen without a pedagogical agent. The personalized voice narration and the personalized text delivered the identical message. Lastly, no verbal or text narration was presented. The personalized on-screen text condition was included to separate the personalized principle from the voice principle and the embodiment principle. The no narration condition was included in the study as a control group by excluding supplementary information, thus presenting only expository information. An example screenshot for each of the six conditions is presented in Figures 6 - 11.

Figure 7. Condition B: Image and personalized on-screen text

Figure 8. Condition C: Image and no personalized narration or text

Figure 9. Condition D: No image and personalized human narration presented by an animated pedagogical agent
Dependent variables

Dependent variables for the study included cognitive load, situational interest, motivation, and achievement.

Cognitive load

Perceived cognitive load was measured by using a single item student self-rating scale developed by Paas and van Merriënboer (1994). Although subjective rating scales are criticized not to provide information regarding which of the three types of cognitive load originated the reported mental effort (Bruïken, Seufert, & Pass, 2010), previous studies shown that this scale is valid, reliable, and sensitive to relatively small differences in cognitive load (Gimino, 2002; Paas, van Merriënboer, & Adam, 1994). In this study, the item asked the participants to use a nine-point Likert-type scale to identify the amount of mental effort they invested to study the instructional material. The cognitive load measures ranged from very, very low mental effort to very, very high mental effort.

Situational interest

The situational interest was measured from three aspects: arousal, involvement, and attention. In order to measure the arousal level, five items from the Activation-Deactivation Adjective Check List (AD-ACL) were used (Thayer, 1986). For example, “I felt active at the moment while I was studying.” The reliability was .91. In order to measure participants’ involvement, two dimensions (intensity and persistence) were considered based on Reynolds'
distinction (Reynolds, 1992). The intensity dimension is measured by self-report items, and the persistence dimension is measured by means of recording the subjects’ participating times (Schiefele & Krapp, 1996). In this study, the intensity dimension was assessed by two items “I was completely caught up in what I was studying”, and “When learning from the material, I was concentrated”. The reliability was .73. Attention was measured as part of situational interest because its’ conceptual proximity to situational interest. Attention deals with the questions such as how to stimulate and sustain a learner’s attention by using novel approaches, creating paradoxes, and using variations in presentation style (Keller, 2010). In order to measure participants’ attention level, twelve attention sub scale from Keller’s Instructional Material Motivation Survey (IMMS) was employed (Keller, 1993). For example, “I found something interesting at the beginning of this instructional material that got my attention.” The response reliability was .83.

**Motivation**

Students’ motivation toward the instructional material was measured in three components of motivation: Relevance, Confidence, and Satisfaction based on the IMMS developed by Keller (1993). The responses ranged from one to five on a Likert scale with nine relevance component items, nine confidence component items, and six satisfaction component items. The reliability of IMMS based on Cronbach’s alpha for each subscale was Relevance: .81, Confidence: .90, and Satisfaction: .92.

**Achievement**

Student achievement was measured on two levels: (1) recall test and (2) comprehension test. A recall test was designed to assess students’ ability to recall as many keywords as possible from the instructional material on intellectual property. The comprehension of the instructional material was measured by a comprehension test that was designed to assess students' ability to select the correct information by applying what they learned from the instructional material, without further access to the material. Items included in this study were six true/false items, three multiple-choice items, and one open-ended question. True/false items were constructed to test students’ ability to remember the factual knowledge about a patent, a trademark, and a copyright correctly. Multiple choice items were constructed to test students’ ability to compare and evaluate related ideas and concepts of a patent, a trademark, and a copyright. One of the items was an open-ended question that referred to the relationship among intellectual property, a patent, a trademark, and a copyright. The purpose of the open-ended question was to assess students’ general understanding of the main topic “Intellectual property.” An example of the true/false item is “A trademark is any sign that includes words, logos, colors, slogans, three-dimensional shapes, but not sounds and gestures.” (T/F). An example of the multiple choice items was “Which one of the followings is not an advantage of registering trademark? (1) Notice to the public of the registrant’s claim of ownership of the mark. (2) A legal presumption of ownership nationwide. (3) The exclusive right to use the mark on or in connection with the goods or services set forth in the registration, and (4). Economic rewards for creator’s efforts.” Lastly, an open-ended question asked “Please describe in writing how the concepts of intellectual property, copyright, trademark, and patent are related each other.”

**Procedure**

When participants logged on the computer, they were guided to the multimedia instructional material on the topic of “Intellectual Property.” And they were asked a series of questions about their demographics, level of individual interest based, and prior learning experience with intellectual property. Then each participant was presented the instructional material corresponding to his/her treatment condition and told to begin the material. They were not allowed to take notes or refer to other resources. At the end of the study, participants were asked to respond to questionnaires for dependent measures.
Data analysis

The study was designed as a 2 × 3 factorial design. The variables included the use of images (presence vs. absence) and the source of narration (pedagogical agent, text without agent, no narration). First, preliminary data analyses were conducted to detect problematic observations and to assess violations of the assumptions for statistical procedures. In a primary data analysis, the main effect of two independent variables was conducted for four dependent variables: (1) perceived cognitive load, (2) situational interest score, (3) achievement score, and (4) motivation score. The significance level for all the analyses was set at α < .05. Bonferroni adjustment was made when multiple comparisons are performed (Field, 2013).

Results

Prior to the main data analysis, the equivalence of treatment conditions in terms of pre-interest (individual interest) was verified. The level of pre-interest was measured using five feeling-related interest items and four value-related interest items based on Schiefele’s definition of individual interest (1991, 1999). The reliability of each measure was .74 and .75 respectively in this study. The result of a one way between-groups multivariate analysis of variance showed no significant differences among the six conditions. Therefore, the level of pre-interest (individual interest) was confirmed equivalent among the six conditions in this study. Additionally, a missing value analysis, a case analysis, and a detection of violations of assumptions for the main dependent variables were conducted. The descriptive statistics for all dependent variables are presented in Table 5 according to the six conditions.

Table 5. Descriptive statistics for the dependent variables

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Measures</th>
<th>Conditions</th>
<th>Image</th>
<th>No image</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Agent (n = 31)</td>
<td>Text (n = 16)</td>
<td>No (n = 17)</td>
</tr>
<tr>
<td>Cognitive load a</td>
<td>Cognitive M</td>
<td>1.87</td>
<td>3.06</td>
<td>2.41</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.02</td>
<td>1.44</td>
<td>1.37</td>
</tr>
<tr>
<td>Situational interest b</td>
<td>Arousal M</td>
<td>2.21</td>
<td>1.56</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.61</td>
<td>.57</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>Involvement M</td>
<td>2.58</td>
<td>2.10</td>
<td>2.44</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.03</td>
<td>.84</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>Attention M</td>
<td>3.47</td>
<td>2.94</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.62</td>
<td>.46</td>
<td>.56</td>
</tr>
<tr>
<td>Motivation c</td>
<td>Relevance M</td>
<td>3.87</td>
<td>3.66</td>
<td>3.28</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.69</td>
<td>.35</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>Confidence M</td>
<td>3.38</td>
<td>2.93</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.53</td>
<td>.50</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>Satisfaction M</td>
<td>2.60</td>
<td>2.10</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.01</td>
<td>.58</td>
<td>.55</td>
</tr>
<tr>
<td>Achievement</td>
<td>Recall test M</td>
<td>7.01</td>
<td>5.25</td>
<td>5.47</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.04</td>
<td>1.53</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>Comprehension test M</td>
<td>5.71</td>
<td>5.13</td>
<td>5.94</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.55</td>
<td>1.50</td>
<td>1.56</td>
</tr>
</tbody>
</table>

Note. a Possible range for cognitive load (1-9); b Possible range for learning interest (1-5); c Possible range for attitude (1-5); d Possible score range for recall test: Minimum: 2, Maximum: 9; e Possible range for comprehension test (0-10).
RQ1. What is the effect of multimedia design principles using social cues on cognitive load in pedagogical agent multimedia learning?

A 2 x 3 between-groups ANOVA on cognitive load measure indicated no significant effect for images, $F(1,121)=2.32$, $p = .13$. However, there was a significant main effect of the source of narrations, $F(2, 121)=4.88$, $p < 0.01$, $\eta^2=.08$. Further post-hoc analysis using a Bonferroni adjusted alpha level of 0.016 revealed that perceived cognitive load was significantly lower when human voice narration was presented by a pedagogical agent ($M = 2.14, SD = 1.24$) than when the narration ($M = 2.94, SD = 1.49$) was not presented. There was no significant difference found between the on-screen text narration condition ($M = 2.84, SD = 1.51$) and the no narration condition although the cognitive load was higher when no narration was presented than the on-screen text narration was presented.

RQ2. What is the effect of multimedia design principles using social cues on situational interest in pedagogical agent multimedia learning?

A factorial MANOVA indicated that there was no statistically significant difference of situational interest score between learners presented with images and learners not presented with images, Wilks’ Lambda = .940, $F(3,119) = 2.54$, $p = .06$. A second factorial MANOVA revealed that there was a statistically significant difference of situational interest score among the three narration conditions, Wilks’ Lambda = .792, $F(6,238) = 4.902$, $p < .001$, $\eta^2 = .11$. Follow-up ANOVA indicated that significant differences occurred in arousal, $F(2,121) = 7.58$, $p = 0.001$, $\eta^2 = .11$, and attention, $F(2,121) = 10.68$, $p < .001$, $\eta^2 = .15$.

Tukey HSD follow-up procedure indicated that arousal score for the pedagogical agent with human voice condition ($M = 2.19, SD = .62$) was significantly higher than both the on-screen text condition ($M = 1.76, SD = .60$) and the no narration condition ($M = 1.77, SD = .60$). For attention, the result indicated that the attention score for the pedagogical agent with human voice condition ($M = 3.30, SD = .61$) was significantly higher than both the on-screen text condition ($M = 2.90, SD = .47$) and the no narration condition ($M = 2.76, SD = .66$).

RQ3. What is the effect of multimedia design principles using social cues on motivation in pedagogical agent multimedia learning?

A factorial MANOVA indicated that there was no statistically significant difference of motivation score between learners presented with images and learners not presented with images, Wilks’ Lambda = .985, $F(3,119) = .599$, $p = .62$. However, a factorial MANOVA revealed that there was a statistically significant difference of learners’ motivation score among the three narration conditions, Wilks’ Lambda = .789, $F(6,238) = 4.99$, $p < .01$, $\eta^2 = .11$. Follow-up ANOVA indicated that significant differences occurred in relevance, $F(2,121) = 8.74$, $p < .001$, $\eta^2 = .13$, and confidence, $F(2,121) = 6.19$, $p < .01$, $\eta^2 = .09$. Tukey HSD follow-up procedure indicated that relevance scores for the pedagogical agent with human voice condition ($M = 3.73, SD = .71$) and the on-screen text condition ($M = 3.68, SD = .46$) were significantly higher than the no narration condition ($M = 3.21, SD = .56$) respectively. For confidence, the result indicated that confidence score for the pedagogical agent with human voice condition ($M = 3.25, SD = .59$) was significantly higher than the no narration condition ($M = 2.82, SD = .67$). However, there was no significant difference found between the on-screen text condition ($M = 2.94, SD = .60$) and the no narration condition.

RQ4. What is the effect of multimedia design principles using social cues on achievement in pedagogical agent multimedia learning?

A 2 x 3 between-groups ANOVA on recall test scores revealed no significant effect for images, $F (1,121)=.001$, $p = .98$, but a significant effect for the source of narration, $F(2,121)= 3.56$, $p < 0.05$, $\eta^2 = .06$. However, further post-hoc analysis using a Bonferroni adjusted alpha level of 0.016 revealed no significant difference among the three narration conditions. Another 2 x 3 between-groups ANOVA on comprehension test scores revealed that there was no
significant effect for images, $F(1,121) = .25, p = .62$, and no significant effect for the source of narration, $F(2,121) = .55, p = .58$.

**Discussion**

This study investigated the effects of multimedia design principles using social cues on learners’ perceived cognitive load, situational interest, achievement, and motivation in pedagogical agent multimedia learning. Overall, the results indicated that using multimedia design principles with social cues increases situational interest and motivation in terms of relevance and confidence. Although the presence of images did not significantly affect the two dependent measures, the source of narration did influence them.

First, it should be noted that the overall cognitive load from all conditions were relatively low (the highest was 3.47 and the lowest was 1.87), which suggests that the material content was not too much difficult for participants to understand thus low intrinsic cognitive load. Yet students’ perceived cognitive load was significantly different between the study conditions. When a personalized human voice narration was presented by a pedagogical agent along with images, students reported the lowest cognitive load whereas students reported the highest cognitive load when no images or narrations were presented (see Figure 12). But, when the personalized text is presented on-screen along with images, students perceived higher cognitive load than when no images were presented. Therefore, we can conclude that the personalization principle has to be applied to the pedagogical agent delivering human voice narration to increase the effectiveness of the multimedia design principles using social cues.

Second, the result of situational interest score analysis shows that learners reported scored significantly higher arousal and attention scores when narrations were presented by a pedagogical agent. The finding indicates that the pedagogical agent delivering human voice narration triggered students to be more interested in the learning material than other conditions. Situational interest is triggered as a result of short-term changes in affective and cognitive processing (Hidi & Renninger, 2006) while being involved in learning activities. The follow-up correlation analysis found a significantly negative correlation between perceived cognitive load and situational interest, $r = -.417, n = 127, p < 0.001$. It implies that highly triggered and maintained situational interest may have reduced perceived cognitive load in pedagogical agent multimedia learning and thus increase generative cognitive processing.

Third, relevance was higher when narrations were delivered either by pedagogical agent’s human voice or on-screen text. It is likely that students were able to link the presented concepts to the real examples provided in the narrations. Confidence was higher only when pedagogical agents delivered the narrations. The results suggest that pedagogical agents delivering narrations are effective in improving students’ perception of relevance and confidence to the learning material. The follow-up correlation analysis showed a significantly negative correlations between perceived cognitive load and relevance, $r = -.259, n = 127, p < 0.01$ and also between perceived cognitive load and confidence,
$r = -0.326$, $n = 127$, $p < 0.001$. It implies that the personalization principle whether used in human voice or in on-screen text can be effective to reduce perceived cognitive load in pedagogical agent multimedia learning and increase generative cognitive processing.

Although multimedia design principles with social cues were found effective to increase situational interest and motivation when applied with a pedagogical agent, no significant differences were found for the recall test and the comprehension test. The reason can be attributed to the four-phases of interest development (Hidi, & Renninger, 2006). Situation interest has to be triggered and maintained before being internalized to well-developed individual interest. In this study, learners showed higher arousal and attention to the learning material, but it is possible that the interest was triggered only by the supplementary information, and the triggered interest might have not been transferred to learning the main expository information. In future research, the triggering point of situational interest in different multimedia elements has to be examined so increase in situational interest can be attributed to the entire learning material, not part of it.

The findings of this study have several implications. First, the results of this study illuminate the importance of using multimedia principles with social cues to increase generative cognitive processing in pedagogical agent multimedia learning. Especially, the social cues can help promote interest and motivation to support concept learning rather than damaging learning. Second, the important finding is the negative correlations between cognitive load and situational interest or motivation. Future work is needed to establish a direct causal effect relationship between the variables.

**Conclusion**

The current study examined four multimedia design principles with social cues to promote generative processing in pedagogical agent multimedia learning. Future research needs to be conducted to verify the study findings. First, as this study mainly focused on applying the multimedia design principles with social cues to designing voice narration by a pedagogical agent and on-text narration, a no-social cue condition also needs to be considered in the future. For example, personalized human voice narrations and personalized machine voice narrations can be compared in different presentation modalities to build comprehensive understanding of the voice principle, and how it affects perceived cognitive load, interest, and motivation differently. Second, this study used a single subject rating method to measure the overall mental effort invested to learning the material, whereas future study would benefit from more contemporary instrument to measure different types of cognitive load separately (see Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013) and examine the relationship between motivation and each type of the cognitive loads. Third, the goal of the learning material was concept learning on the topic of intellectual property. This might have caused a little or no interaction between the elements in the learning material, thus reduced the perceived cognitive load. Future study should extend the scope of material content to other subject areas with high level of element interaction such as math or science so that students can be fully engaged in generative cognitive processing to solve a problem. Lastly, this study confirmed the equivalent level of individual interest across the conditions before the main analyses, yet the possible effect of prior knowledge was not considered. Additional research is needed to include a pretest to measure learners’ prior knowledge and investigate how learners’ prior knowledge and the multimedia design principles interact in multimedia learning.

**References**


228


The Beast of Aggregating Cognitive Load Measures in Technology-Based Learning

Jimmie Leppink* and Jeroen J. G. van Merriënboer
School of Health Professions Education, Maastricht University, The Netherlands // jimmie.leppink@maastrichtuniversity.nl // j.vanmerrienboer@maastrichtuniversity.nl

*Corresponding author

ABSTRACT
An increasing part of cognitive load research in technology-based learning includes a component of repeated measurements, that is: participants are measured two or more times on the same performance, mental effort or other variable of interest. In many cases, researchers aggregate scores obtained from repeated measurements to one single sum or average score per participant and use these aggregated scores in subsequent analysis. This paper demonstrates some dangers of this commonly encountered aggregation approach and presents two comprehensive alternatives: Split-plot analysis of variance (ANOVA) and more flexible two-level regression analysis. The core message of this paper is that the application of the aggregation approach can seriously distort our view of effects and relations of interest and should therefore not be used in cognitive load research. Multilevel analysis of repeated measurements data can account for various features of the data and constitutes a best practice.

Keywords
Technology-based learning, Medical education, Repeated measurements, Split-plot design, Multilevel analysis

Introduction
An increasing part of cognitive load research in technology-based learning, and cognitive load research in general, includes a component of repeated measurements. Then, participants – frequently students, employees, managers or clients – and in statistical language more commonly referred to as subjects are measured repeatedly (i.e., at least two times) on the same variable(s) of interest. Be it performance, mental effort invested in a cognitive activity or some other variable, repeated measurements data have one nice feature: they enable researchers to separate so-called within-subjects variance from between-subjects variance. When we have a sample of individual students perform an exam only once, we can distinguish between participants, for some students perform better than others do. In such a context, however, we cannot account for variation within participants, since we have only one measurement per student. In educational research, we are usually not interested exclusively in distinguishing between participants; we are interested in within-participants changes related to learning and development as well, and for the latter we need studies that include a component of repeated measurements.

Recent calls for repeated measurements of cognitive load
Recently, two series of well-designed randomized controlled experiments (Schmeck, Opfermann, Van Gog, Paas, & Leutner, 2015; Van Gog, Kirschner, Kester, & Paas, 2012) provided evidence for the statement that studies where learners have to perform a series of tasks, it is better to measure a characteristic of interest – for instance mental effort invested in a cognitive activity (Paas, 1992) – after each task (i.e., repeatedly) than once respectively. A main finding of both series of experiments was that single retrospective (i.e., delayed) ratings were significantly higher than the average of ratings obtained after each task. This probably happens because delayed ratings are mainly influenced by the relatively more complex problems (Schmeck, et al., 2015).

In another experiment (Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013), students solved problems on conditional and joint probabilities in two modes: in an explanation of six lines of text, and in formula notation. In one condition, students first studied the textual explanation and then the formula explanation, while in the other condition the order was reversed. Various cognitive load measures were included in this experiment (Ayres, 2006; Cierniak, Scheiter, & Gerjets, 2009; Leppink, et al., 2013; Paas, 1992; Salomon, 1984). However, instead of administering these measures at the end of the full order of two formats, the measures were administered after each format. This enabled researchers to not only test for differences between formats but also test for format order effects. One finding that has implications for instructional design is that students who are confronted with the
formula format before the textual explanation tend to experience higher extraneous cognitive load from the formula format than their peers who study exactly the same materials on joint and conditional probabilities in reversed order. The latter implication could not have resulted from aggregation of extraneous cognitive load scores across formats.

Repeated measurements data enable researchers to separate between-participants and within-participants variance. Unfortunately, many researchers have failed to appreciate this nice feature of repeated measurements data and have aggregated scores from repeated measurements to one sum or average score per participant (e.g., Ayres, 2006; Corbalan, Kester, & Van Merriënboer, 2008; Hoffman, 2012; Koriat, Nussinson, & Ackerman, 2014; Kostons, Van Gog, & Paas, 2012; Van Loon, De Bruin, Van Gog, & Van Merriënboer, 2013). Aggregating repeated measurements, we wipe out all within-participants variance and this can result in serious distortions of our views of effects and relations of interest (Leppink, 2015). Figure 1 provides an example of what can happen when studying the correlation between two quantitative variables that have been measured repeatedly but for which repeated measurements have been aggregated.

In the example depicted in Figure 1, four students (small number for the sake of illustration) perform five tasks in counterbalanced order and rate the invested mental effort after each of five tasks. Each task yields one mental effort rating and one quantitative performance score. When we aggregate the scores for both measures – or compute a Pearson’s correlation coefficient for the twenty (i.e., not aggregated) observations in the scatterplot – we find a positive correlation between mental effort and performance (left part of Figure 1). Thus, we would conclude – in this specific context – that students who report higher mental effort tend to display better performance as well. However, when taking the repeated measurements structure of the data into account, we find a negative correlation between mental effort and performance: for a given student, we would conclude that a task that requires more mental effort appears to yield somewhat lower performance (right part of Figure 1). The latter message has much more meaning in many educational settings, and provides a better interpretation of the available data, than the former message. Note that the example in Figure 1 is hypothetical; it could well be that an aggregation approach yields a negative correlation between mental effort and performance while taking the repeated measurements structure of the data into account results in a positive or no (linear) correlation at all. The point is that the conclusions might become different when the repeated measurements structure is taken into account when analyzing the data.

Monitoring of learning and mental effort

One context in which the comparison between aggregating scores vs. treating repeated measurements as such is important is the context of self-regulated learning. Self-regulated learning can be defined as an active, constructive, metacognitive process (i.e., metacognition is cognition about cognition, in this case: learning; Flavell, 1979). Important metacognitive processes within self-regulated learning are monitoring and control (Zimmerman & Schunk, 2001). The term monitoring is used to refer to the thoughts learners have about their cognition, and learners respond to the environment (e.g., adapt their behavior) based on these metacognitive thoughts. The latter is termed control (Nelson & Narens, 1990).
Self-regulated learning can be studied at three levels: at item level, at task or topic level, and at task-sequence level. Self-regulated learning at item level is about learners’ monitoring of how well they have learned smaller pieces of information (e.g., word pairs, small passages of text), which affects how long they engage in (re)studying it (Metcalfe, 2009; Thiede & Dunlosky, 1999). Self-regulated learning at task or topic level involves learners’ monitoring of their understanding of information on a particular topic, which determines whether they continue studying one piece of information or move on to another (Azevedo & Cromley, 2004; Azevedo, Moos, Greene, Winters, Cromley, 2008). Finally, self-regulated learning at the level of task sequence is about learners’ monitoring of how well they performed a learning task after completing it. This is also referred to as “self-assessment” to distinguish it from monitoring during task performance and is used by learners to select next suitable tasks from an environment comprising tasks at different levels of complexity and with different levels of instructional guidance (Corbalan, et al., 2008; Kostons, et al., 2012).

Effectiveness of self-regulated learning in which learners can choose their own learning tasks greatly depends on the accuracy of students’ self-assessment and task selection (Kostons, et al., 2012). Unfortunately, accurate self-assessment is very difficult for learners, and especially novices, as it appears to require some domain expertise and knowledge of assessment criteria (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Inaccuracies in task selection can arise when self-assessment is inaccurate, but even when self-assessment is accurate students frequently do not know how to select suitable tasks. When task selection is inaccurate, this will cause learners to work on tasks that do not match their knowledge level—meaning selected tasks are either too complex or too easy or tasks that have an inappropriate amount of instructional guidance and therefore impose a suboptimal cognitive load on students’ minds—and consequently, students will not learn from engaging in self-regulated learning.

In a recent study (Kostons, et al., 2012), students worked on a series of genetics tasks, self-rated task performance on a scale from zero (minimum) to five (maximum) and mental effort on a scale from one (minimum) to nine (maximum) after each task, to use these ratings for selecting the next task (which out of five complexity levels and how much instructional support). Data obtained from such a study hold very interesting information about experienced mental effort from self-selected tasks, can as such provide feedback to learners and teachers about the appropriateness of a given task in terms of complexity and instructional support for a given student at a given stage, and provides a wealth of information for researchers interested in ways to improve the accuracy of such ratings to eventually improve the accuracy of task selection. Unfortunately, instead of approaching the data with flexible multilevel analysis, all ratings and performance measures were aggregated across all tasks performed by students, thereby not differentiating between ratings and performance on easier vs. more complex tasks or between ratings and performance on tasks with more vs. less instructional support. The same has been done in other studies in this context (e.g., Corbalan, et al., 2008; Kostons, Van Gog, & Paas, 2010; Kostons, et al., 2012; Salden, Paas, & Van Merriënboer, 2006).

Supporting the acquisition of self-regulated learning skills is of crucial importance in contemporary education for two reasons. Firstly, such skills are critical to effective learning in schools and other environments. Secondly, a rapidly changing society in which learners must be ready to develop new knowledge and skills autonomously and continuously does a great appeal on self-regulated learning skills. Continuing developments in technology-based education increasingly create room for both the application and research of self-regulated learning.

In fact, all three levels at which self-regulated learning takes place involve repeated measurements with smaller or larger intervals of time in between. Unfortunately, in virtually all studies published in this context, scores obtained from these repeated measurements are aggregated to sum or average scores. Yet, especially in a context of self-regulated learning, using repeated measurements as such appears to make much more sense than using aggregated scores. Suppose, for instance, that a student reports to have invested a rather high mental effort on a task on which this student performs rather poorly. It makes more sense to have the student (choose to) perform a somewhat easier task, and have the student base subsequent task selection on changes in reported mental effort and performance from task to task.

**Scales for assessing cognitive load**

Subjective rating scales have been used intensively to measure or estimate cognitive load experienced by learners, since the introduction of a nine-point one-dimensional mental effort rating scale by Paas (1992). Mental effort has
been defined as the cognitive capacity allocated to deal with the demands imposed by a cognitive task and as such reflects the actual and overall cognitive load (Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Paas & Van Merriënboer, 1994a). The overall cognitive load is assumed to be the sum of intrinsic cognitive load and extraneous cognitive load (Kalyuga, 2011; Sweller, 2010; Sweller, Ayres, & Kalyuga, 2011). Intrinsic cognitive load is a direct function of the complexity of a task on the one hand and the student’s prior knowledge on the other hand (Sweller, 1994). More advanced students have more prior knowledge and can therefore be expected to experience a lower intrinsic cognitive load than less advanced students confronted with the same task. Extraneous cognitive load is due to features in the way in which information (e.g., instruction, explanation) is presented as a result of which the student has to invest mental effort that does not contribute to learning or task performance and is therefore extraneous to learning (Sweller & Chandler, 1994; Sweller, Chandler, Tierney, & Cooper, 1990). An optimal intrinsic cognitive load can be achieved through a proper match between the complexity of tasks and the learner’s prior knowledge, while extraneous cognitive load should be minimized. Doing so, we can stimulate students to allocate still available working memory resources to deal with the intrinsic cognitive load and, as such, engage in learning. In a traditional version of cognitive load theory, the latter was referred to as a third category of cognitive load called germane cognitive load (Sweller, Van Merriënboer, & Paas, 1998). However, a recent reconceptualization of the categories of cognitive load resulted in a two-factor intrinsic/extraneous framework in which germane cognitive load is viewed as working memory resources allocated to dealing with intrinsic cognitive load (Kalyuga, 2011; Sweller, 2010; Sweller, et al., 2011). A series of recent studies have provided support for the possibility to distinguish between intrinsic and extraneous cognitive load using a questionnaire comprising three to four statement items with subjective rating scales for each category of load (Leppink, et al., 2013; Leppink, Paas, Van Gog, Van der Vleuten, & Van Merriënboer, 2014).

An advantage of the multi-item questionnaire developed by Leppink et al. (2013, 2014) to the aforementioned single-item mental effort rating scale introduced by Paas (1992) is that it enables researchers to distinguish between two fundamentally different categories of cognitive load; a single-item score simply cannot do that. It is statistically speaking impossible to measure or estimate to what extent a difference on the single-item mental effort rating scale can be attributed to differences in intrinsic or extraneous cognitive load without being absolute sure through solid experimental design that one type of cognitive load is kept constant across conditions. Only in the case that an experimental study fully succeeded in keeping intrinsic cognitive load constant across conditions one might want to interpret changes in mental effort as changes in extraneous cognitive load. However, even in that case one might find researchers debate how one can be sure that intrinsic cognitive load is constant across conditions without having any measure available as a check. Using a multi-item instrument, we can study whether we can use that multitude of items (i.e., directly observed or manifest variables) to measure or estimate one or more latent variables (i.e., variables that cannot be observed directly), in this context types of cognitive load (e.g., Leppink, et al., 2013, 2014). With a single item, that exercise is impossible.

The distinction between intrinsic and extraneous cognitive load is relevant, because an increase in cognitive load does not always hamper learning. When extraneous cognitive load is low, slight increases in intrinsic cognitive load may in fact stimulate learning in that there is more to learn from a task of some complexity than from a task that is very easy given a student’s prior knowledge and therefore imposes a very low intrinsic cognitive load on the student’s mind. In the latter case, there is virtually no need to allocate working memory resources to deal with intrinsic cognitive load, because there is hardly any intrinsic cognitive load to deal with in the first place.

The most recent version of the questionnaire developed by Leppink et al. (2014) comprises four items for measuring intrinsic cognitive load and four items for measuring extraneous cognitive load. Evidence from the same study suggests, however, that three items for each of the two categories – as reported by Leppink et al. (2013) – can be sufficient. That makes six items for intrinsic and extraneous cognitive load together. A questionnaire of such a size is easy to administer in a single measurement or in a limited number of, for instance, five repeated measurements. A disadvantage of a multi-item questionnaire like this one, however, is that it is more difficult to administer in a study that for instance has twenty repeated measurements. If students (or other individuals) have to learn or perform five tasks, rating six questions after each task may still not be asked too much from the students. However, having to rate six questions after each of twenty tasks is likely to be perceived as a tedious exercise perhaps even before the students have completed ten tasks. This may result in over-consistent rating or other undesirable responding that of course has little to do with actual intrinsic or extraneous cognitive load.
To unite the best of the six-item intrinsic/extraneous cognitive load questionnaire and the single-item mental effort rating scale, a solution in the case of as many as twenty tasks could be to ask students to provide a mental effort rating after each task and only complete the intrinsic/extraneous cognitive load questionnaire after a block of four or five tasks. This way, repeated measurements are available for both mental effort and the two categories of cognitive load, while the risk of undesirable responding is minimized.

**Randomized controlled experiments**

A major strength of cognitive load theory is that the guidelines and implications coming forth from this theory are generally based on randomized controlled experiments. One might argue that the term “randomized” is redundant here, given that the nature of experiments is that participants are randomly assigned to the categories of an independent variable; without such randomization, we would be dealing with a quasi-experiment. We nevertheless add the term “randomized” because the term “experiment” is sometimes encountered in situations when a study reported was actually a quasi-experiment. Further, the term “controlled” is used because in an experiment, an independent variable is systematically manipulated and, thus, there is a kind of control. Well-designed randomized controlled experiments are of great importance in educational research, because they allow causal conclusions with regard to effectiveness of instruction or some other treatment on learning or performance. Randomized controlled experiments enable us to study the effectiveness of instructional methods before they are implemented in education rather than afterwards.

**Example**

Let us consider the following example. The advent of technology has contributed enormously to possibilities to study the human brain. The human brain is a very complex structure and studying it is a challenge for many people, including students. Suppose that some researchers have developed a new supportive tool to actually help medical students in their study of the anatomy of the human brain. The human brain is a large three-dimensional structure, which is very difficult to represent in two-dimensional pictures. Two-dimensional pictures of (parts of) the brain are already complex study objects, but having to mentally build a three-dimensional picture of the brain or of a brain region is something that can require students to invest a high mental effort. The researchers have developed the new supportive tool to provide specific guidance to students while learning about the anatomy of the human brain. It is therefore expected to reduce the mental effort that must be invested by the students.

To study the effectiveness of this new tool, the researchers assign 100 students to either the experimental treatment condition ($n = 50$) or control condition ($n = 50$). Students in the control condition perform five learning tasks in logical order in a traditional way, while students in the experimental treatment condition perform the same learning tasks in the same order with help of the new supportive tool. In both conditions, students rate their mental effort on a visual analog scale from 0 (minimum) to 100 (maximum) after each of five tasks. The researchers then aggregate these five repeated measurements to one average rating per student and compare the two groups in terms of average rating. Since the new tool, used in the experimental treatment condition, is expected to reduce mental effort invested by the students, the researchers expect the average rating to be significantly lower in the experimental treatment condition than in the control condition.

**Split plots**

The type of study design in the aforementioned example is also known as split-plot design. This term stems from agricultural experiments in which split plots of land received different treatments and were monitored or measured across time (Fisher, 1925). Likewise, in the current example study, students are allocated to different treatment conditions and are measured repeatedly on the same variable of interest (Howell, 2012), more specifically, five times on mental effort.

The approach chosen by the hypothetical researchers in the example study constrains them to be satisfied with a comparison of the difference between the two conditions in rating averaged over the five tasks. However, what if:
• use of the supportive tool helps to reduce mental effort only after a number of tasks, meaning that it does not affect mental effort in the first couple of tasks;
• use of the supportive tool helps to reduce mental effort in the first couple of tasks (i.e., an initial stage) but its effect decreases and eventually vanishes;
• use of the supportive tool leads to an elevated mental effort in the first couple of tasks but decreases mental effort in subsequent tasks; or
• use of the supportive tool results in a decreased mental effort in the first couple of tasks but elevates mental effort in subsequent tasks?

These four scenarios point at a so-called interaction between condition and task: the effect of condition (i.e., supportive tool) depends on (i.e., is moderated by) task or, in other words, the effect of condition is not constant across tasks. Each of these four scenarios can have important implications for self-regulated learning and/or guidance needed in (self-regulated) learning, yet none of these four interaction scenarios can be studied by the aggregation approach. In fact, the aggregation approach ignores all possible interactions between condition and task (i.e., measurement occasion) and assumes (among others) that the effect of condition is the same across tasks, we may find no (significant) differences between conditions, while a repeated measurements approach may indicate how the effect of condition depends on task.

A not rarely heard argument from experimental researchers is that we do not need to test for the aforementioned interaction if we do not have a hypothesis about it beforehand or that there is only a need for that if a hypothesis about a particular main effect (in this context: a difference between conditions across tasks or time) is not supported by the data. However, interpretations of main effects (and interpretations of interaction effects likewise) are meaningful only if crucial assumptions underlying the tools used for analysis are met.

Non-interaction is such an assumption, even in for instance analysis of covariance (ANCOVA) or factorial analysis of variance (ANOVA) on single-measurement data when reporting main effects (e.g., Field, 2013; Howell, 2012; Huitema, 2011). This and other assumptions (e.g., normality, homogeneity of variance, independence of observations) should be examined not only if study results seem to not provide support for hypotheses but also if the results at first appear to provide such support. One of the potential outcomes of such a critical assessment of assumptions in the latter may be that a powerful conclusion about support for a hypothesis cannot be made given that non-interaction (or another important assumption) is not met and must therefore be accounted for through some change(s) in the statistical model. One of the problems with the aggregation approach is that non-interaction is automatically assumed and cannot be tested. This is a problem because condition by time or task interaction is quite common (Howell, 2012; Leppink, 2015; Tan, 2008), also in cognitive load research (Leppink, et al., 2013), and implies that what we have thought of as a main effect (i.e., the difference between conditions being constant across time or tasks) may vary quite a bit across time or tasks.

In the case of a so-called “ordinal” interaction (i.e., non-parallel lines that do not cross), main effects can strictly speaking be interpreted. If for instance an experimental treatment group and a control group are measured in terms of task performance three times, the average change across measurement occasions is different for the two groups but the average task performance is higher in one and the same group at each of the three occasions, it is fair to speak of a main effect of group even if the interaction (i.e., difference in change between groups) is statistically significant. However, even in such a case, a significant interaction term comprises information that is not present in main effects. If for instance the experimental treatment administered prior to the three measurements needs some time to have an effect on task performance, we may well see that the groups do not yet differ significantly on the first occasion but that group differences are larger and statistically significant at subsequent measurement occasions. Main effects do not comprise that information. Yet, when aggregating repeated measurements to one single score, the only thing we can interpret is a main effect. Finally, in the case of disordinal interactions, that is: when lines do cross, one should not interpret main effects at all (Howell, 2012), because which group performs better depends on another variable, here measurement occasion. This paper discusses an example of a disordinal interaction: the experimental treatment group starts off with slightly higher average mental effort (albeit it non-significantly) but at some point the difference reverses to have a significantly lower mental effort in the experimental group on the final two occasions.
**Internal consistency vs. test-retest reliability**

When interpreting aggregated scores, researchers sometimes refer to a “satisfactory” or “high” internal consistency of the scores on the tasks aggregated (e.g., Ayres, 2006; Corbalan, et al., 2008; Hoffman, 2012; Paas, 1992; Paas et al., 2003; Paas & Van Merriënboer, 1994b). Internal consistency is largely about the extent to which items assumed to measure the same characteristic yield similar scores. Hordes of researchers compute and report Cronbach’s alpha coefficient (Cortina, 1993; Cronbach, 1951) as an estimator of internal consistency, which is about the extent to which items assumed to measure the same characteristic yield similar scores. This approach of computing internal consistency (although Cronbach’s alpha is not always the best choice; Peters, 2014) and using factor scores through aggregation or through factor analysis generally makes good sense for multiple items measuring the same characteristic at one single point of time (e.g., three items for intrinsic or extraneous cognitive load, Leppink et al., 2013; 2014). However, applying that approach to single-item repeated measurements is a tricky enterprise. The reason for the latter is that, in a context like the one in the example study, two sources of variance resonate and are perfectly confounded in any estimate of internal consistency: variance due to error in the (here: mental effort) rating and variance due to task differences. In other words, we cannot know to what extent a Cronbach’s alpha value reflects (a lack of) reliability in measurement in this specific context or variation due to differences between tasks. Further, in the following, it becomes clear that even a high Cronbach’s alpha value does not imply absence of any of the aforementioned types of interaction.

A more appropriate perspective on reliability in studies of this kind is that of test-retest reliability, as estimates of the latter provide an indication of the extent to which a measure provides consistent results for the same participants at different measurement occasions. One would expect that participants who experience above average mental effort at the first measurement occasion are to some extent also more likely to experience above average mental effort at subsequent measurement occasions, especially in studies where the kind of treatment (either experimental or control group) is the same across a range of tasks of the same kind.

**Method**

The remainder of this paper demonstrates the aforementioned problem of ignoring potential condition by task interaction by aggregating repeated measurements in a randomized experiment and presents two comprehensive alternatives accounting for the repeated measurements. For educational purposes, data from this example study were simulated, and a detailed overview of the simulation procedure is available from the authors.

The advantage of a simulation study is that the outcomes of the study are known and as such it enables a comparison of strengths and weaknesses of various methods of analysis. The data were simulated such that, at the level of population of (possible) students, the supportive tool has no effect on mental effort for the first three tasks but results in a substantially lower mental effort compared to the control condition at subsequent tasks (in this case: the fourth and fifth task). Software programs such as SAS, SPSS, STATA, Mplus, and R provide facilities for dealing with repeated measurements data as done in the remainder of this paper. Since among educational researchers SPSS appears to be used much more than any of the other programs mentioned, it is perhaps interesting to mention that SPSS v21 was used for analysis in this case.

**Approach (1): Aggregation**

As mentioned previously, students in both conditions rate their mental effort on a visual analog scale from 0 to 100 after each of five tasks. The researchers aggregate these five repeated measurements to one average rating per student, and compare the two conditions in terms of average rating, expecting the average rating to be significantly lower in the experimental treatment condition. This comes down to a two-sample t-test or one-way ANOVA on the aggregated scores.
Approach (2): Split-plot ANOVA

Split-plot ANOVA – which is in fact a repeated measures ANOVA that includes at least one between-participants factor (e.g., treatment in this example study) – is a well-known tool among experimental psychologists (Howell, 2012), who for instance compared two or more conditions in a pretest posttest (and follow-up) control-group design in one study or another, the difference between conditions being a different treatment (or, in the case of the control condition, no treatment at all) between pretest and posttest. Split-plot ANOVA provides a valid tool for analysis if there is no missing data due to some students omitting one rating or another and certain assumptions with regard to variances of scores on tasks and correlations between task scores are realistic. If there is no missing data, split-plot ANOVA works fine if the variance of scores is not too different across tasks and not too different between conditions and if the correlation between scores is not that much different across pairs of tasks. Otherwise, the third approach – which is more flexible – may provide a slightly better alternative.

Approach (3): Two-level regression

In fact, split-plot ANOVA is a special type of a two-level regression model or, also called, mixed-effects model. The term mixed effects is used whenever an analysis involves one or more fixed effects and one or more random effects. The purpose of a study like this one is to generalize findings to a larger population of (possible) students, and we assume that the students in our study form a random sample from a population that has a particular and preferably Normal (i.e., bell-shaped) distribution. In other words, we treat “student” as a random effect. We can estimate student-specific random effects because we have repeated measurements from the same students. Treatment, however, is a fixed effect; we are interested in the specific comparison of experimental treatment and control condition, and we do not consider these two treatments as a random sample of a universe of possible treatments to which we generalize. Likewise, the repeated measurements are in this context treated as fixed effect; the interest lies in differences between conditions on these specific tasks and, to some extent, in differences between tasks as well. In the type of two-level regression we are discussing here, student and measurement occasion (1, 2, 3, 4, 5) are treated as hierarchical levels with student being the upper level. Split-plot ANOVA is a special type of two-level regression model in that it is somewhat restrictive in assumptions with regard to variances and correlations as mentioned before, assumptions that can be relaxed in a more flexible two-level regression approach as is demonstrated in the last paragraph of the Results section.

Results

Table 1 presents descriptive statistics per task per condition, and Figure 2 provides a graphical representation of the mean rating per task per condition.

<table>
<thead>
<tr>
<th>Rating (0-100)</th>
<th>Experimental treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>60.80 (5.00)</td>
<td>59.72 (5.28)</td>
</tr>
<tr>
<td>Second</td>
<td>59.38 (6.36)</td>
<td>57.74 (6.65)</td>
</tr>
<tr>
<td>Third</td>
<td>55.78 (6.91)</td>
<td>55.38 (7.71)</td>
</tr>
<tr>
<td>Fourth</td>
<td>46.36 (8.50)</td>
<td>52.82 (8.53)</td>
</tr>
<tr>
<td>Fifth</td>
<td>40.12 (10.94)</td>
<td>48.62 (9.95)</td>
</tr>
</tbody>
</table>

Table 1 and Figure 2 indicate that the average mental effort is slightly higher in the experimental treatment condition in the initial stage (i.e., the first three tasks) but then drops and is lower than in the control condition after the fourth and fifth task.
Researchers following the commonly encountered aggregation approach typically omit the step as done in Table 1 and Figure 2. Instead, they report that the five ratings yield a Cronbach’s alpha of 0.85 for the two conditions together and even slightly higher when computed for experimental treatment condition (0.87) and control condition (0.87) separately. These high Cronbach’s alpha values are then taken as justification for aggregating ratings obtained after different tasks to one average rating to then proceed as follows: the average rating in the experimental treatment condition of 52.51 (SD = 6.35) is slightly lower than the average rating of 54.88 (SD = 6.34) in the control condition. However, the difference is rather small, $\eta^2 = 0.034$ (values around 0.01, 0.06, and 0.14 indicate small, medium, and large effects, respectively), and not statistically significant at the conventional 0.05 significance level, $F(1, 98) = 3.493, p = 0.065$. What follows in the case of such an outcome is an attempt to find an explanation for the unexpected result and/or a plea to replicate the study in some slightly different context, without considering the possibility of treatment-by-task interaction.

**Approach (2): Split-plot ANOVA**

Table 1 indicates that for a given task, the two conditions have similar standard deviations. Levene’s test per task reveals that the standard deviations of the two conditions are not statistically different in any of the tasks (with $p$-values ranging from 0.377 for the third task to 0.907 for the fourth task). Further, Box’s test reveals that the covariance matrix for the five repeated measurements is not significantly different between the two conditions ($p = 0.863$). Finally, Mauchly’s test indicates that the standard error for pairwise comparisons differs significantly across tasks ($p < 0.001$), which is not a surprise because Table 1 indicates that in both conditions the standard deviation around the mean rating is about twice as large in the fifth task than in the first task. We can to some extent correct for the latter by using an adjusted test, here Huynh-Feldt (Field, 2013), which in this specific case adjusts all degrees of freedom ($df$) by a factor 0.500 meaning that all degrees of freedom become twice as small. While this adjustment is somewhat conservative in that it comes at the cost of loss of some statistical power for detecting effects of interest (i.e., somewhat increased probability of Type II error), the adjustment provides some protection against an inflation of the probability of finding significant differences while in fact there is no effect (i.e., Type I error probability). An alternative is to use one of the multivariate tests (e.g., Pillai’s Trace), which do not require (to adjust for departure from) equal standard errors.
The first thing we test is the treatment-by-task or split-plot interaction: $F(2,001, 196.142) = 21.889, p < 0.001, \eta^2 = 0.183$. It turns out that the non-parallel lines (i.e., interaction) displayed by (Table 1 and) Figure 1 is statistically significant and quite large. What we can do now is so-called simple effects analysis, that is: we perform a two-sample $t$-test or one-way ANOVA for the difference in average rating between conditions per task. If the interaction is not statistically significant, we do not need to perform simple effects analysis for then the difference in average rating between conditions does not differ significantly across tasks. Split-plot ANOVA provides the same test outcome as found in the aggregation approach for the difference in average rating over all five tasks together, and it is – provided that other assumptions are met, more about this follows when presenting the third approach – then interpretable. However, in the case of significant interaction, as is the case right now, it makes much more sense to inspect the difference between conditions per task.

Simple effects analysis reveals no significant difference between conditions on the first task, $F(1, 98) = 1.103, p = 0.296 (\eta^2 = 0.011)$, no significant difference on the second task, $F(1, 98) = 1.587, p = 0.211 (\eta^2 = 0.016)$, no significant difference on the third task, $F(1, 98) = 0.075, p = 0.785 (\eta^2 = 0.001)$, a statistically significant difference on the fourth task, $F(1, 98) = 14.382, p < 0.001 (\eta^2 = 0.128)$, and a statistically significant difference on the fifth task, $F(1, 98) = 16.532, p < 0.001 (\eta^2 = 0.144)$. In other words, differences between conditions appear small and rather meaningless on the first three tasks, whereas after the fourth and fifth task the average mental effort is quite a bit lower in the experimental treatment condition. Finally, one note of caution on the reported $\eta^2$-values: even though in articles and in some statistical software output the term eta-squared or $\eta^2$ is used, the $\eta^2$-values reported are in fact partial eta-squared values (Field, 2013), because they reflect the proportion of variance explained by the effect under consideration that is not explained by other effects in the model (i.e., the unique contribution of the effect under consideration).

### Approach (3): Two-level regression

It is clear that the split-plot ANOVA approach is more appropriate than the aggregation approach. However, split-plot ANOVA is quite restrictive in terms of correlations between repeated measurements: they should be more or less equal. In randomized experiments in which tasks are counterbalanced such that the tasks are provided to different participants in different order, that assumption may be feasible. In the current experiment, one and the same logical order is used. Further, split-plot ANOVA tends to be somewhat less feasible if there are considerable differences in standard deviation across tasks. Table 1 indicates that this is the case; the standard deviation increases throughout and is about twice as large after the fifth task than after the first task. A more flexible two-level regression approach allows for taking eventual departures from equal correlations and eventual departures from equal standard deviations into account. Elaborating on all possible ways to do so could easily fill a full paper on itself (Tan, 2008; Verbeke & Molenberghs, 2000). In the remainder of this section, we focus on the most likely candidate models for this specific context.

Given that the tasks are provided in one specific order (because this specific order appears logical in this study context), it is to be expected that ratings on subsequent tasks are more correlated than ratings from tasks further away from each other. In other words, task pairs 1-2, 2-3, 3-4, and 4-5 (i.e., nearest task pairs) can be expected to have the highest correlation. The correlation for task pairs 1-3, 2-4, and 3-5 (i.e., second-nearest task pairs) could be somewhat lower than for the aforementioned four task pairs but still somewhat higher than the correlation for task pairs 1-4 and 2-5 (i.e., third-nearest task pairs). Finally, task pair 1-5 (i.e., fourth-nearest task pair) is to be expected to yield the lowest correlation, because of the largest distance between the first and fifth task. If the four nearest task pairs have (more or less) one common correlation, the three second-nearest task pairs have (more or less) one common correlation (that is different from that of the nearest task pairs), and the two third-nearest task pairs have (more or less) one common correlation (that is different from the correlation of nearest task pairs and different from the correlation of the second-nearest task pairs), we are dealing with a so-called Toeplitz structure (Bareiss, 1969; Leppink, et al., 2013). A more restrictive version of this structure applies when the nearest task pairs have correlation $r$, the second-nearest task pairs have correlation $r^2$ (i.e., $r$ times $r$) the third-nearest task pairs have correlation $r^3$ (i.e., $r$ times $r$ times $r$), and the final task pair (i.e., 1-5) has a correlation of $r^4$ (i.e., $r$ times $r$ times $r$ times $r$). This is a so-called first-order autoregressive (AR1) structure. Given that the correlation between any pair of tasks (including nearest tasks) is usually smaller than one, AR1 implies that there is a predictable decline in correlations with increasing space between tasks. AR1 is a likely candidate in for instance growth studies (Tan, 2008). If AR1 fits the data well, its advantage over Toeplitz is that we have a more parsimonious model, that is: we use fewer degrees of freedom.
freedom for accounting for this correlation structure and keep these degrees of freedom for a more powerful test on effects of interest (here: treatment). However, if AR1 turns out too restrictive, Toeplitz is a common alternative candidate.

Commonly used criteria for comparing the fit of AR1, Toeplitz and other structures are Akaike’s information criterion (AIC) (Akaike, 1973), Schwarz’s Bayesian information criterion (BIC) (Schwarz, 1978) and the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, & Van der Linde, 2002). All these criteria have in common that they indicate the extent to which the model deviates from the data, and smaller values mean less deviance or, in other words, better fit. The (relative) parsimony of models under consideration is taken into account, since all these criteria penalize for the number of degrees of freedom consumed by the models under consideration. Further, the BIC has a penalty for the sample size ($N$). In the context of repeated measurements data, a danger of the latter is that $N$ can be defined in different ways since it can refer to observations at different levels (Fitzmaurice, Laird, & Ware, 2004; Hedeker & Gibbons, 2006; Weakliem, 1999). Since this distinction is not made in the BIC, the BIC may sometimes favor (too) simple models. Finally, although there is some indication that the DIC can yield good performance in the context of multilevel analysis (Gelman & Hill, 2007), this criterion still needs more research and for now it (perhaps for that reason) is not included in some standard statistical packages like SPSS.

Let us compare six models in terms of the AIC:

- Model 1: all task pairs have the same correlation (i.e., compound symmetry) and the standard deviation is equal across tasks. This is more or less conform traditional split-plot ANOVA;
- Model 2: all task pairs have the same correlation (i.e., compound symmetry) but the standard deviation is not equal across tasks;
- Model 3: AR1 structure and the standard deviation is equal across tasks;
- Model 4: AR1 structure and the standard deviation is not equal across tasks;
- Model 5: Toeplitz structure and the standard deviation is equal across tasks; and
- Model 6: Toeplitz structure and the standard deviation is not equal across tasks.

<p>| Table 2. Outcomes of Model 6 (Toeplitz and unequal standard deviations) (based on simulated data) |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>$B$ ($SE$)</th>
<th>$df$</th>
<th>$t$-value</th>
<th>$p$-value</th>
<th>Confidence interval (95%)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>59.72 (0.83)</td>
<td>96.589</td>
<td>71.923</td>
<td>&lt; 0.001</td>
<td>58.07</td>
<td>61.37</td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td>57.74 (1.03)</td>
<td>103.263</td>
<td>56.010</td>
<td>&lt; 0.001</td>
<td>55.70</td>
<td>59.78</td>
<td></td>
</tr>
<tr>
<td>Task 3</td>
<td>55.38 (1.03)</td>
<td>111.801</td>
<td>53.778</td>
<td>&lt; 0.001</td>
<td>53.34</td>
<td>57.42</td>
<td></td>
</tr>
<tr>
<td>Task 4</td>
<td>52.82 (1.09)</td>
<td>128.729</td>
<td>48.245</td>
<td>&lt; 0.001</td>
<td>50.65</td>
<td>54.99</td>
<td></td>
</tr>
<tr>
<td>Task 5</td>
<td>48.62 (1.33)</td>
<td>134.968</td>
<td>36.438</td>
<td>&lt; 0.001</td>
<td>45.98</td>
<td>51.26</td>
<td></td>
</tr>
<tr>
<td>Task 1 by Treat</td>
<td>1.08 (1.17)</td>
<td>96.589</td>
<td>0.920</td>
<td>0.360</td>
<td>-1.25</td>
<td>3.41</td>
<td></td>
</tr>
<tr>
<td>Task 2 by Treat</td>
<td>1.64 (1.46)</td>
<td>103.263</td>
<td>1.125</td>
<td>0.263</td>
<td>-1.25</td>
<td>4.53</td>
<td></td>
</tr>
<tr>
<td>Task 3 by Treat</td>
<td>0.40 (1.46)</td>
<td>111.801</td>
<td>0.275</td>
<td>0.784</td>
<td>-2.49</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td>Task 4 by Treat</td>
<td>-6.46 (1.55)</td>
<td>128.729</td>
<td>-4.172</td>
<td>&lt; 0.001</td>
<td>-9.52</td>
<td>-3.40</td>
<td></td>
</tr>
<tr>
<td>Task 5 by Treat</td>
<td>-8.50 (1.89)</td>
<td>134.968</td>
<td>-4.504</td>
<td>&lt; 0.001</td>
<td>-12.23</td>
<td>-4.77</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effect</th>
<th>Variance ($SE$)</th>
<th>Wald $z$</th>
<th>$p$-value</th>
<th>Confidence interval (95%)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable $^2$</td>
<td>Task 1</td>
<td>34.47 (4.96)</td>
<td>6.949</td>
<td>&lt; 0.001</td>
<td>26.00</td>
<td>45.71</td>
</tr>
<tr>
<td>Task 2</td>
<td>53.14 (7.09)</td>
<td>7.185</td>
<td>&lt; 0.001</td>
<td>40.45</td>
<td>69.80</td>
<td></td>
</tr>
<tr>
<td>Task 3</td>
<td>53.02 (7.09)</td>
<td>7.477</td>
<td>&lt; 0.001</td>
<td>40.80</td>
<td>68.92</td>
<td></td>
</tr>
<tr>
<td>Task 4</td>
<td>59.93 (7.07)</td>
<td>8.023</td>
<td>&lt; 0.001</td>
<td>46.94</td>
<td>76.52</td>
<td></td>
</tr>
<tr>
<td>Task 5</td>
<td>89.02 (10.84)</td>
<td>8.215</td>
<td>&lt; 0.001</td>
<td>70.13</td>
<td>113.01</td>
<td></td>
</tr>
</tbody>
</table>

| Correlation | Nearest | 0.79 (0.02) | 34.861 | < 0.001 | 0.74 | 0.83 |
| Second-nearest | 0.58 (0.05) | 12.882 | < 0.001 | 0.49 | 0.66 |
| Third-nearest | 0.38 (0.07) | 5.522 | < 0.001 | 0.24 | 0.51 |
| Task pair 1-5 | 0.20 (0.09) | 2.149 | 0.032 | 0.01 | 0.38 |

Note. $^1$ Treatment is coded as follows: 0 = control condition, 1 = experimental treatment condition; $^2$ The variance is the standard deviation squared.
The AIC values are 3226.018 (Model 1), 3167.441 (Model 2), 3054.532 (Model 3), 3022.713 (Model 4), 3049.593 (Model 5), and 3021.908 (Model 6). Generally, the models that account for the fact that (as indicated in Table 1) the standard deviations vary quite a bit across tasks (i.e., Model 2, Model 4, and Model 6) perform better than the models that assume the standard deviation to be equal across tasks (i.e., Model 1, Model 3, and Model 5). Model 4 and Model 6 perform best, indicating that the assumption of the correlation being (more or less) equal across tasks is not realistic, and there appears to be a slight preference for the latter. Table 2 therefore presents the outcomes of Model 6.

Given the coding (i.e., 0 = control condition, 1 = experimental treatment condition), Task 1 represents the average rating after the first task in the control condition, and Task 1 by Treat represents the difference in average rating after the first task between the two conditions. In other words, the average rating is slightly higher in the experimental treatment condition after the first task, as can be seen in Table 1 and Figure 1. Task 2 represents average rating after the second task in the control condition, and Task 2 by Treat represents the difference in average rating after the second task between the two conditions. The same interpretation holds for the other three tasks, so Task 5 represents the average rating after the fifth task in the control condition while Task 5 by Treat represents the difference in average rating after the fifth task between the two conditions. In other words, we see that the two conditions do not really diverge in the first three tasks but on the fourth and fifth task, you see a clear difference between the two conditions. Since we view a decrease in mental effort as something positive in this context, the difference between the two conditions after the fourth and fifth task is in favor of the experimental treatment condition and thus in favor of the supportive tool.

Discussion

Some readers, especially the ones rooted in split-plot ANOVA, may wonder why go for a more complex two-level model if split-plot ANOVA leads to the same conclusion with regard to the treatment of interest. Indeed, both approaches beat the aggregation approach for an obvious reason: instead of wiping out all within-student variance, we account for it and conclude that the treatment has the expected effect only after a number of tasks. While the aggregation approach results in the mere conclusion that the supportive tool has a rather small and not statistically significant effect, both alternatives indicate that there is little to no reason to assume any treatment effect for the first three tasks but after that the expected positive effect of the supportive tool becomes evident.

If there is no missing data and departures from equal correlations and equal standard deviations are not that large, split-plot ANOVA generally provides a quick and efficient way to draw valid conclusions and present the findings. However, in the case of more severe departures from equal correlations and equal standard deviations, split-plot ANOVA tends to use inappropriate standard errors for particular hypothesis tests. Moreover, in the case of missing data, split-plot ANOVA tends to produce biased estimates, because participants who accidentally omit one rating or another are deleted completely in the split-plot ANOVA approach even if four of the five ratings are available.

Repeated measurements and required sample size

Treating repeated measurements data as such also has benefits for the statistical power or probability of detecting an effect in a random sample if that effect exists at population level. For instance, using G*Power (Buchner, Erdfelder, Faul, & Lang, 2009) – a program that enables researchers to calculate among others statistical power given statistical significance level, sample size, expected effect size, and for more complex studies some other factors or the required sample size given a desired statistical power, statistical significance level, sample size, and expected effect size – reminds us that, given a medium size effect and $\alpha = 0.05$ in a standard two-tailed test, to obtain a statistical power of 0.80 in a two-group comparison of means from a single measurement occasion, we need a sample size of $n = 64$ per group. Making the same kind of main effects comparison when treating two repeated measurements with a correlation of 0.50 (which is often reasonable: Hedeker & Gibbons, 2006) as such instead of aggregating these to a single mean score per group, we need a total sample size of $n = 49$ per group for that same statistical power. Further, when dealing with five measurements with an average between-measurements correlation of 0.50 (as in our example study), the sample size required for a power of 0.80 is only about $n = 39$ per group. To obtain a power of 0.80 for the
group-by-measurement interaction in such a study including five measurements, we need only about \( n = 11 \) per group.

### Learning groups

Another type of study in which the validity of split-plot ANOVA is limited, is when individual participants are nested within, for instance, learning groups in which participants interact and cooperate for a particular period of time. Just like repeated measurements typically induce a within-participant between-measurements correlation, participants interacting and cooperating in learning groups tends to induce a within-group between-participants correlation. In the example study in this paper, students receive treatments individually. If treatment is applied at the level of small (e.g., problem-based) learning groups, it is recommendable to account for that in the analysis stage. With ANOVA we cannot do that, with multilevel analysis we can (Snijders & Bosker, 2011). For power or required sample size calculations it is always better to consult a statistician in such a case, as sample sizes at both the level of learning groups and the level of participants within learning groups depend on the magnitude of both fixed and learning-group-level random effects and it is therefore hard to give general guidelines that hold for most cases, except for the following.

The number of learning groups is more important than large numbers of individuals per learning group (Hox, 2010). Current-day software can provide reasonably accurate estimates for learning-group level variances when including as few as six to twelve learning groups (Browne & Draper, 2000), but of course estimates and their standard errors become more accurate with increasing sample sizes at both levels (i.e., learning groups and individuals). When interested mainly in experimental treatment effects, dealing with learning groups of size fifteen or smaller may be no problem (e.g., thirty learning groups of fifteen students each; Leppink, 2015), and when learning group effects are on the large side, multilevel analysis may be feasible even if learning groups comprise as few as two learners only (i.e., study in pairs; Leppink, Broers, Imbos, Van der Vleuten, & Berger, 2012). However, when dealing with effects of smaller size or interested in cross-level interactions, one may want to consider a 30/30 or 50/20 rule of thumb (Hox, 2010), where the first number refers to the number of groups and the latter to the number of individuals per group.

### Departures from assumptions

Of course, any statistical model is based on assumptions. Some assumptions may be more realistic than others, and – in terms of impact of assumption departures on model outcomes – some assumptions are more important than others. For instance, most tests involving comparisons of means are fairly robust against some skewness. The assumption of homogeneity of variances is an important one for ANOVA models but can be relaxed in more flexible regression models. Further, in studies in which participants are nested within learning groups and also in quite some studies involving repeated measurements, random intercepts and/or random slopes are specified. These random intercepts and slopes are learning-group-level effects in the case participants are nested within learning groups, participant-level effects if individual participants are measured repeatedly, and in studies where individual participants nested within learning groups are measured repeatedly we expectedly have a combination of both (Leppink, 2015). The assumption typically made with regard to these random intercepts and slopes is that they are normally distributed around the fixed intercept and slope, respectively. For the kind of piecewise linear models used in this paper, some departure from these normality assumptions is generally not a serious problem, but for non-linear models and categorical response variable models things are trickier (McCulloch & Neuhaus, 2011). This leads to the final assumption, namely that – in the models discussed in this paper – we assume the response variables of interest to be of interval or ratio level of measurement. When dealing with learning-group nesting or repeated-measurements studies in which response variables of interest are binary or ordinal, generalized linear mixed models or generalized estimating equations (Hedeker & Gibbons, 2006; Molenberghs & Verbeke, 2005) should be used, and the latter also offer an alternative if interval/ratio level response variables show more severe departures from normality or homogeneity of variances. Finally, when dealing with a response variable that is a count (e.g., the number of typing errors in a task), a special case of the generalized linear mixed model, the Poisson model, can be used (Hox, 2010).
Conclusion

The message that we need to treat repeated measurement scores as they are, namely as scores obtained from the same participants measured repeatedly, applies to any study in educational research that includes repeated measurements data. No matter what level of self-regulated learning we are talking about – item level, task/topic level or task sequence level – we run a risk of drawing inappropriate conclusions whenever we aggregate repeated measurements data. This holds not only for mental effort or for the specific context of self-regulated learning. The message applies to any study that deals with the repeated measurement of performance, motivation, satisfaction, pain or other variable of interest. As Figure 1 indicates, caution with aggregation is also needed when examining relations between quantitative variables. This distinction also has relevance in experiments in the context of self-regulated learning, in which students select and perform a series of for instance eight tasks of either of five possible difficulty levels (1, 2, 3, 4, 5), rate their mental effort after each task, and receive instructional support (i.e., experimental treatment condition) or do not receive instructional support (i.e., control condition) in this process of task selection. The instructional support may then affect not only performance and mental effort rating, but even on which tasks the performance and mental ratings are based. We may see considerable differences between as well as within conditions (between and within students) in terms of performance, mental effort rating, and task selection pattern. The current practice is to aggregate performance across tasks, aggregate mental effort ratings across tasks, and aggregate difficulty level across tasks to some kind of “average” difficulty, perform ANOVA or regression on aggregated scores while ignoring all meaningful information in the repeated measurements. To estimate the effects of treatment, task difficulty level, and perhaps other variables of interest on performance, mental effort, intrinsic or extraneous cognitive load, we need multilevel analysis.

The core message of this paper is that the approach of aggregating repeated measurements data to single (sum or) average scores can distort our view of effects and relations of interest to a substantial extent and should therefore not be used in cognitive load research. Multilevel analysis of repeated measurements data that can account for various features of the data constitutes a best practice.

References


Characterization of Educational Resources in e-Learning Systems Using an Educational Metadata Profile

Georgia Solomou1*, Christos Pierrakeas1,2 and Achilles Kameas1

1Hellenic Open University (HOU), Patras, Greece // 2Technological Educational Institute (TEI) of Western Greece, Patras, Greece // solomou@eap.gr // pierrakeas@eap.gr // kameas@eap.gr

*Corresponding author

(Submitted March 27, 2014; Revised March 23, 2015; Accepted May 1, 2015)

ABSTRACT

The ability to effectively administrate educational resources in terms of accessibility, reusability and interoperability lies in the adoption of an appropriate metadata schema, able of adequately describing them. A considerable number of different educational metadata schemas can be found in literature, with the IEEE LOM being the most widely known; however, it is often the case where it cannot fully accommodate the characteristics of several types of educational resources, that’s why application profiles have been proposed. Each metadata standard and application profile usually comes with a different (either less or more semantically enriched) binding, thus allowing the retrieval and dissemination of resources with varying degrees of effectiveness. In this work, we propose an application profile of the IEEE LOM standard having special focus on distance learning material, while being generic enough so as to be applicable to any educational material and application. We then present an ontology model for this profile that aims to improve the potential discovery and retrieval of educational resources within intelligent e-learning environments.

Keywords

Learning objects, Ontologies, Metadata application profile distance education, IEEE LOM

Introduction

Metadata are “machine-readable information about electronic resources or other things” (Berners-Lee, 1997) and are used to describe the features of a resource, thus making easier its management and retrieval. A set of metadata elements combined so as to serve a specific purpose, constitute a metadata schema.

Although the adoption of a single metadata standard would assure reusability of resources and interoperability among applications, there exists no metadata schema yet, appropriate to fulfil the requirements and needs of every application. Some schemas focus on technical metadata, other on educational metadata while some other on more specialized elements. When existing approaches are not sufficient enough to cover the special requirements of an institution or organization, the use of application profiles is suggested. According to Heery & Patel (2000), an application profile is an aggregation of metadata elements selected from one or more different schemas and combined into a new compound schema.

Particularly, in the case of educational resources, the set of metadata used to describe their characteristics, should be able to capture their educational and pedagogical aspects. Therefore, apart from author, title or type – fields that are common in all metadata schemas – an educational metadata schema should also include information regarding the resource’s particular learning type, its intended end users, the instructional context and many more.

A kind of educational resource that is increasingly used by Educational Institutions in recent years is the Learning Object (LO). According to Nikolopoulos, Solomou, Pierrakeas & Kameas (2012) LOs are pieces of educational material that directly correlate the knowledge they convey with specific objectives (learning outcomes) of the learning process. But although LOs constitute a common trend in organizing educational material and have been utilized by many modern e-learning systems (Schreurs & Al-Zoubi, 2007), they cannot be used effectively because there exists no metadata schema capable of capturing all of their characteristics.

This insufficiency becomes even greater in the case of LOs that are designed for use in the context of distance learning courses, where the proper handling and dissemination of LOs is crucial for the success of the learning process, because in most cases, contrary to what happens in a classroom, no human tutor would be continuously available to monitor students’ path or progress through the educational process.
Hellenic Open University (HOU) is a Higher Education Institute specialized in distance and lifelong learning that the last two years seeks to re-organize its material and to provide its students with advanced services for delivering knowledge. Such services require the consumption of adequately characterized LOs, using a metadata schema that is capable of capturing as many as possible of their pedagogical aspects and especially those considered to be important according to distance learning principles. To the best of our knowledge, no such a schema or application profile exists, able to satisfy these requirements.

Consequently, through this work we propose an application profile of the IEEE LOM standard with special focus on the field of distance learning. After reviewing existing approaches for describing educational resources, as well as several binding methods (section Background), we move on to the presentation of our proposed Educational Metadata Profile (EMP), in section EMP: New Elements and Modifications. Its ontological binding is given in the subsequent section (The EMP ontology), whereas section Evaluation of the EMP ontology presents an evaluation of this ontology model, through its application for characterizing real LO instances. Conclusions follow, in our last section.

Background

In literature, several metadata standards and profiles have been proposed, each serving different purposes and needs. A few examples of them are mentioned in the following subsection.

Educational metadata standards and profiles

The IEEE Learning Technology Standards Committee (LTSC) has developed a standard for the description of learning material and learning resources, known as IEEE Learning Object Metadata (IEEE LOM) (Hodgins & Duval, 2002). IEEE LOM is without doubt a widespread standard for educational metadata and focuses mainly on the description of educational resources and especially LOs. It includes more than 60 elements classified into 9 categories (General, Life Cycle, Meta-Metadata, Technical, Educational, Rights, Relation, Annotation, Classification), each one of them containing metadata for various aspects of a LO, including its technical characteristics and rights, as well as educational and instructional features.

Several IEEE LOM profiles can be found in literature. Each of them is usually designed so as to accommodate very specific needs of its corresponding organization/creator. ARIADNE (see http://www.ariadne-eu.org/) is one such profile that intends to describe learning material used in secondary and post-secondary education. It is designed with the aim to solve two major problems: (a) the indexing of educational resources (i.e., the creation of the metadata by persons) and (b) the exploitation of metadata by users who look for relevant pedagogical material (that should be as easy and efficient as possible).

Another IEEE LOM profile is IMS Learning Resource Metadata (IMS LRM) (Nilsson, 2001), which constitutes a set of specifications for learning resources. It includes elements useful for the description of learning resources, while the specifications address issues like content packaging, question and test interoperability, learning design and simple sequencing. IMS LRM adopts all of the LOM’s categories and elements.

Two more LOM application profiles, that were created in order to describe resources locally, are CanCore (Canadian Core) (http://cancore.athabascau.ca/en/) and UK LOM Core (http://zope.cetis.ac.uk/profiles/uklomcore/). CanCore, used mainly in Canada, simplifies LOM and at the same time maximizes interoperability between different projects. UK LOM Core, designed for United Kingdom educational system, intends to provide guidelines to those who desire to create, use and apply metadata.

The Sharable Content Object Reference Model (SCORM) is a reference model which controls how the learning content is organized, described and linked with Learning Management Systems. SCORM allows the extension of LOM, thus enabling organizations to add new elements and enhance the existing controlled vocabularies.
On the other hand, the Dublin Core Metadata Initiative (DCMI) (DCMI, 2012) has been developed by organizations so as to aid the sharing of any kind of generic web resources. Its initial version, the Dublin Core Metadata Element Set (DCMES) (DCMI, 2012), known as Dublin Core (DC), consists of 15 elements. The subsequent Qualified Dublin Core (QDC) (DCMI Usage Board, 2012) extends DC with 7 new elements. However, even this enriched version of the DC schema is unable to capture the pedagogical aspects of an educational resource.

GEM (Gateway to Educational Materials) (http://dublincore.org/groups/education/GEM-Study.html) is a DCMI profile that is education-oriented. It is an RDF metadata vocabulary, designed for the description of educational resources and, apart from the DCMI elements, encompasses several additional educational-specific properties. GEM includes controlled vocabularies for the different levels of end users, evaluation methods and tools as well as the types of resources.

Although all of the aforementioned standards and profiles manage to capture some of the most important pedagogical characteristics of educational resources, they become weak when applied to distance education, because, in order to be effective, the educational material designed to serve distance learning courses has to conform to specific requirements regarding its content, layout, structure, technical properties, etc. (Lockwood, 2013). This material assumes the role of the tutor of a distance learner, so it has to be explicitly correlated to learning outcomes and be available in various formats, in order to be able to cover various learning styles. So, even the very promising IEEE LOM standard, despite its generality, still fails to specify some of these important educational aspects of learning resources, like for example the expected learning outcomes of a course, the direct correlation among learning outcomes and educational material, as well as the different (learning) types of the latter that are used within a distance learning course.

Binding of metadata elements

The elements of a metadata schema are usually handled as SQL tables, text files, HTML meta-tags and so on Nejdl et al. (2002). Such a technical realization of the abstract model of a metadata schema in a specific format is called a binding. Structured formats for binding the elements of a metadata schema are XML, XML Schema, RDF, RDF Schema and OWL.

XML provides a surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents. XML Schema is a language for restricting the structure of XML documents and also extends XML with datatypes. The Resource Description Framework (RDF) (RDF Working Group, 2014) is a datamodel for objects (“resources”) and relations between them and provides simple semantics which in turn can be represented using XML syntax. RDF Schema is a vocabulary for describing properties and classes of RDF resources, with the necessary semantics for generalization – hierarchies of such properties and classes. Even though RDF is intended for representing knowledge, it lacks reasoning abilities; RDF does not support making inferences or deductions. Therefore, a much more expressive framework is required, so that metadata can be meaningfully encoded.

Ontologies, expressed in OWL (OWL Working Group, 2014), are the pillar of the Semantic Web and provide the ability to represent any domain of interest in a more structured way. OWL provides the vocabulary for describing properties and classes (relations between classes, cardinality, equality, richer typing of properties, characteristics of properties and enumerated classes) and poses constraints about what statements the user can declare.

For IEEE LOM, XML and RDF bindings are available, implemented by the IMS Global Learning Consortium (Nilsson, 2001; Nilsson, Palmér & Brase, 2003). The usage of XML for the LO metadata expression facilitates the indexing process and the retrieval of annotated learning resources. However, this format seems to be not sufficient enough to address the limitation of text-based searching, since XML does not provide the meaning of the described structures. RDF attempts to overcome the problem by adding semantics to each metadata element. Therefore, the description of LOM elements via RDF facilitates their integration into e-learning systems which nowadays are dominated by Semantic Web technologies.

A great number of web-based educational systems exist that embed ontological models in their implementation (Al-Khalifa & Hugh, 2006). These ontological models can reflect various aspects of an e-learning system, such as
student profiles and knowledge domains. The integration of an educational resource’s components to such systems requires semantically richer representation (http://dublincore.org/groups/education/GEM-Study.html). Therefore, many research groups have attempted to annotate semantically the educational metadata. Some representative examples are ALOCoM, SCORM, and an ontology based on the ACM Computer Classification System (ACM CCS).

More specifically, the ALOCoM ontology (http://hmdb.cs.kuleuven.be/alocom/) consists of three parts: the ALOCoM Core ontology, the ALOCoM Content Structure and the ALOCoM Content Type. It has been used in the LO repository of the ARIADNE Foundation as a format to store well-structured and easy-to-reuse LOs. It is a generic content model that defines a framework for LOs and their components. That is, it describes the structural elements of a LO, it focuses on potential pedagogical roles of content units and defines concepts that describe the components common to any type of a LO.

In Yang, Chen, Tsai and Chao (2005) a systematic approach called the Visualized Online Authoring Tool model (VOAT) is developed. Using this model, educators and instruction designers can construct SCORM 2004-compliant courses. Through the Ontology-Based Outline Authoring Tool (OBOAT) educators can be guided by the domain ontology and thus more accurately construct an outline of their course, based on their own individual teaching contexts (beliefs, preferences and student characteristics).

Finally, a well-known example of ontology for the computer science domain has been based upon the ACM Computer Classification System (http://www.acm.org/class) and is defined using RDFS (Brace & Nejdl, 2004). This ontology, which is specific to the educational resources purpose and structure, has been used in the Edutella system (Nejdl et al., 2002) to classify LOs in order to improve searching.

All of the aforementioned ontologies reveal an attempt by educational content providers to create richer representations of the metadata elements they use to characterize their resources. Such representations can have many advantages in the learning process, from alleviating the design of a course to improving the discovery of educational resources. In addition, ontologies are fundamental to improve interaction with users and intelligent agent systems and make systems open to external knowledge integration (Wray, Lisse & Beard, 2004).

**EMP: New elements and modifications**

In order to create an educational metadata schema suitable for the efficient characterization of distance learning material, the design principles of distance learning should be taken into account. Such a schema should be flexible and functional, so that it can capture a large number of educational aspects and pedagogical features of educational resources and make them a prominent means in the knowledge discovery and delivery process. In addition, it should be in accordance with the requirements of other structures that manage educational resources, like institutional libraries, which are usually based on a cataloguing standard.

Having these in mind, we propose a new application profile of the IEEE LOM standard that, unlike all other existing profiles, makes provision for the particular features of the distance learning material. We opted for IEEE LOM, due to its relatively wide acceptance in the academic environment and its extensive usage by institutional repositories. To build this profile, we took into account the guidelines provided by CEN/ISSS (Smith, Van Coillie & Duval, 2006), according to which we had to accomplish the following steps:

- identify the specific characteristics of the distance learning material;
- identify which of these characteristics are reflected in the standard, existing elements of the base schema (IEEE LOM);
- modify the base metadata schema according to these specific requirements (extend it with additional, new elements, modify value space and/or data type of existing elements);
- provide a binding.
The proposed Educational Metadata Profile (EMP) adopts a subset of the IEEE LOM element set and augments it with some new attributes in order to represent concepts commonly used in distance education. This new application profile has been designed to be rich enough, so as to effectively describe some of the most important aspects of an educational resource (pedagogical, technical, etc.), but not exceedingly analytic as to become difficult to use.

In what follows (subsection Structure) we give the list of EMP elements and summarize the modifications and additions that we made to the base schema. In the next subsection Mapping to DCMI metadata elements, we provide a mapping of the resulting schema to the widely accepted DC metadata standard.

**Structure**

Table 1 summarizes the structure elements of the proposed EMP. Those that have been directly taken from IEEE LOM are considered to come with associated semantics. All other additions and modification have been categorized into three groups, as follows:

- The first group (Group A) reflects elements that appear with differences in the value space of their controlled vocabulary, as compared to their counterparts in the original IEEE LOM schema.
- The second group (Group B) refers to elements that come with modifications in their definition and data type only.
- Finally, the third group (Group C) includes completely new elements, with no counterparts in the original IEEE LOM schema. These new entries address several important characteristics of the distance learning material.

<table>
<thead>
<tr>
<th>Name</th>
<th># Repetitions</th>
<th>Datatype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifier$^1$</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>Identifier. Catalog</td>
<td>1</td>
<td>CharacterString</td>
</tr>
<tr>
<td>Identifier. Entry</td>
<td>1</td>
<td>CharacterString</td>
</tr>
<tr>
<td>Title</td>
<td>1</td>
<td>LangString</td>
</tr>
<tr>
<td>Language</td>
<td>10</td>
<td>CharacterString</td>
</tr>
<tr>
<td>Description</td>
<td>10</td>
<td>LangString</td>
</tr>
<tr>
<td>Keyword</td>
<td>10</td>
<td>LangString</td>
</tr>
<tr>
<td>Aggregation Level</td>
<td>1</td>
<td>Controlled Vocabulary</td>
</tr>
<tr>
<td><strong>Life Cycle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contribute$^1$</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td>Contribute. Role</td>
<td>1</td>
<td>Controlled Vocabulary</td>
</tr>
<tr>
<td>Contribute. Entity</td>
<td>40</td>
<td>CharacterString</td>
</tr>
<tr>
<td>Contribute: Entity: Affiliation$^{***}$</td>
<td>1</td>
<td>CharacterString</td>
</tr>
<tr>
<td>Contribute. Date</td>
<td>1</td>
<td>DateTime</td>
</tr>
<tr>
<td><strong>Technical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Format$^*$</td>
<td>40</td>
<td>Controlled Vocabulary</td>
</tr>
<tr>
<td>Size</td>
<td>1</td>
<td>Integer</td>
</tr>
<tr>
<td>Requirement$^{**}$</td>
<td>40</td>
<td>LangString</td>
</tr>
<tr>
<td>Duration</td>
<td>1</td>
<td>CharacterString</td>
</tr>
<tr>
<td><strong>Educational</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Resource Type$^*$</td>
<td>8</td>
<td>Controlled Vocabulary</td>
</tr>
<tr>
<td>Intended End User Role</td>
<td>10</td>
<td>Controlled Vocabulary</td>
</tr>
<tr>
<td>Instructional Context$^{***}$</td>
<td>10</td>
<td>Controlled Vocabulary</td>
</tr>
<tr>
<td>Typical Age Range</td>
<td>1</td>
<td>LangString</td>
</tr>
<tr>
<td>Difficulty</td>
<td>1</td>
<td>Controlled Vocabulary</td>
</tr>
<tr>
<td>Typical Learning Time</td>
<td>1</td>
<td>CharacterString</td>
</tr>
<tr>
<td>Learning Outcome$^1$</td>
<td>1000</td>
<td>-</td>
</tr>
<tr>
<td>Learning Outcome. Identifier$^1$</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>Learning Outcome. Identifier.Catalog$^{***}$</td>
<td>1</td>
<td>CharacterString</td>
</tr>
<tr>
<td>Learning Outcome. Identifier.Entry$^{***}$</td>
<td>1</td>
<td>CharacterString</td>
</tr>
</tbody>
</table>
Group A summarizes the elements that come with modifications in their predefined set of accepted values. These modifications reflect our attempt to meet more accurately the specific characteristics of a resource. The elements that have changed in this way are the following: (1) Format of the Technical category, (2) Learning Resource Type of the Educational category and (3) Kind of the Relation category.

More specifically, Format has now a new set of allowable values, as summarized in Table 2. This set is based on the official IANA MIME media types (Freed, Klensin, & Hansen, 2013) and has emerged after considering the characteristics of the educational material that is already used by HOU. Consequently, it was carefully selected so as to be quite broad and able to cover a wide range of technical data types met in a distance education institution.

Table 2. Possible values of the format element (i.e., technical data types of the resource)

<table>
<thead>
<tr>
<th>Element</th>
<th>Value space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>document</td>
</tr>
<tr>
<td></td>
<td>hypertext</td>
</tr>
<tr>
<td>Image</td>
<td>photo</td>
</tr>
<tr>
<td></td>
<td>map</td>
</tr>
<tr>
<td></td>
<td>graph</td>
</tr>
<tr>
<td></td>
<td>image</td>
</tr>
<tr>
<td></td>
<td>presentation</td>
</tr>
<tr>
<td>Streaming Media</td>
<td>audio recording</td>
</tr>
<tr>
<td></td>
<td>animation</td>
</tr>
<tr>
<td></td>
<td>self-running presentation</td>
</tr>
<tr>
<td></td>
<td>webcast</td>
</tr>
<tr>
<td></td>
<td>video</td>
</tr>
<tr>
<td>Application</td>
<td>interactive software</td>
</tr>
<tr>
<td></td>
<td>hypermedia application</td>
</tr>
<tr>
<td></td>
<td>wiki</td>
</tr>
<tr>
<td></td>
<td>presentation</td>
</tr>
</tbody>
</table>

Learning Resource Type comes with substantial modifications in the set of its accepted values. The main problem with its original value space – as given in the IEEE LOM specification – is that it consists of values which express both educational information (e.g., Exercise, Problem Statement, Simulation) and technical information (e.g., Diagram, Figure, Graph) of an educational resource. In our schema, the technical information (form) is captured by the Format element, so Learning Resource Type should normally address only the various types of educational resources according their educational content. In addition to this separation, some important types of educational resources, such as Example, Serious Game, Case Study and Project, which were absent from the original value space, were added.

To this end, we define a completely new list of acceptable values that reflect the most common types of educational material used within distance education courses and incorporate only information related to the instructional
The value space of the Kind element consists of a controlled vocabulary expressing the various kinds of relationships among educational resources and especially LOs. But apart from the default relationship has part and its inverse is part of, two additional types of relationships are needed so as to capture all possible interconnections among LOs:

- supports and its inverse is supported by attempts to correlate a “supportive” LO that contains complementary or prerequisite knowledge, with an LO that has a key role in the learning process (a “core” LO).
- is alternative type correlates two or more LOs that have exactly the same educational content and differ only in their technical format. This is a highly significant relationship, especially in the case of personalized learning because it makes use of the learners’ preferences.

The only element of the IEEE LOM schema that has been adopted with different definition and data type (Group B of modification), is Requirement, which is used to describe any particular need in terms of software or hardware. The data type of Requirement has been altered to LangString. Therefore, the value of this element contains a short description of any special software or hardware requirements, e.g., “The use of Adobe Acrobat Reader required, version 6.xx or newer.” The corresponding element in the original LOM schema required a more detailed description and more fields to be filled. We thus simplify the process of describing the technical requirements of a resource, keeping at the same time all necessary information intact.

<table>
<thead>
<tr>
<th>Learning resource type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Guidelines</td>
<td>Self-Assessment</td>
</tr>
<tr>
<td>Presentation</td>
<td>Multiple Choice Questions</td>
</tr>
<tr>
<td>Demonstration</td>
<td>Open Type Question</td>
</tr>
<tr>
<td>Lecture</td>
<td>Problem Statement</td>
</tr>
<tr>
<td>Definition-Principle-Law</td>
<td>Experiment</td>
</tr>
<tr>
<td>Narrative Text</td>
<td>Serious Game</td>
</tr>
<tr>
<td>Analogy</td>
<td>Exercise</td>
</tr>
<tr>
<td>Example</td>
<td>Multiple Choice Questions</td>
</tr>
<tr>
<td>Activity</td>
<td>Open Type Question</td>
</tr>
<tr>
<td>Case Study</td>
<td>Problem Statement</td>
</tr>
<tr>
<td>Problem Solving</td>
<td></td>
</tr>
<tr>
<td>Text Composition</td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td></td>
</tr>
<tr>
<td>Interactive</td>
<td></td>
</tr>
<tr>
<td>Non Interactive</td>
<td></td>
</tr>
</tbody>
</table>

Finally, we have augmented our schema with some new elements (Group C) that are necessary in order to represent concepts commonly used in distance education.

- Affiliation keeps important information about the life cycle of a resource as it determines the status of the entity that has contributed to the creation and development of this particular resource.
- Learning Outcome is placed under the Educational category and expresses the correlation of an LO with one or more learning outcomes. In particular, for each learning outcome that an LO satisfies, one needs to give a natural language statement, via the Description element, as well as to assign to it an identifier (Identifier:Entry and Identifier:Catalog respectively), according to a specific identification system.
- Instructional Context implies the actual context where the learning process takes place, and can accept values like “distance education,” “face to face learning” and “blended learning.” This is a key element because it reveals the mode of learning for which the particular object is appropriate.

Elements in Group A and C mostly address modifications/additions that render our EMP suitable for adequately describing educational resources used in the context of a distance learning course. New elements have been introduced (Learning Outcome, Instructional Context), additional values for the Learning Resource Type have been
provided – capturing the various types of distance learning material – as well as new relationships that directly correlate learning outcomes and LOs, have been declared.

Mapping to DCMI metadata elements

The EMP described above has particular orientation in education. Nevertheless, to facilitate the potential implementation of our EMP by applications based on the widely used DC schema, we provide a mapping of the EMP to DCMI metadata elements (see Table 4). Apart from those elements that had a direct correspondence to DC (e.g., title, language), the rest of EMP elements are mapped to those DC terms that are closer in meaning.

Table 4. Mapping of the EMP elements to the DC metadata terms

<table>
<thead>
<tr>
<th>IEEE LOM (EMP)</th>
<th>DC metadata terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td></td>
</tr>
<tr>
<td>Identifier.Catalog</td>
<td>isPartOf</td>
</tr>
<tr>
<td>Identifier.Entry</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>title</td>
</tr>
<tr>
<td>Language</td>
<td>language</td>
</tr>
<tr>
<td>Description</td>
<td>description</td>
</tr>
<tr>
<td>Keyword</td>
<td>subject</td>
</tr>
<tr>
<td>Aggregation Level</td>
<td>isPartOf</td>
</tr>
<tr>
<td>Life Cycle</td>
<td></td>
</tr>
<tr>
<td>Contribute.Role</td>
<td>contributor, creator, publisher</td>
</tr>
<tr>
<td>Contribute.Entity</td>
<td>description</td>
</tr>
<tr>
<td>Contribute.Entity.Affiliation</td>
<td>description</td>
</tr>
<tr>
<td>Contribute.Date</td>
<td>date</td>
</tr>
<tr>
<td>Technical</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>format</td>
</tr>
<tr>
<td>Size</td>
<td>extent</td>
</tr>
<tr>
<td>Requirement</td>
<td>requires</td>
</tr>
<tr>
<td>Duration</td>
<td>extent</td>
</tr>
<tr>
<td>Educational</td>
<td></td>
</tr>
<tr>
<td>Learning Resource Type</td>
<td>type</td>
</tr>
<tr>
<td>Intended End User Role</td>
<td>audience</td>
</tr>
<tr>
<td>Context</td>
<td>educationLevel</td>
</tr>
<tr>
<td>Typical Age Range</td>
<td>description</td>
</tr>
<tr>
<td>Difficulty</td>
<td>type</td>
</tr>
<tr>
<td>Typical Learning Time</td>
<td>temporal</td>
</tr>
<tr>
<td>LearningOutcome.Identifier.Catalog</td>
<td>isPartOf</td>
</tr>
<tr>
<td>LearningOutcome.Identifier.Entry</td>
<td>identifier</td>
</tr>
<tr>
<td>LearningOutcome.Identifier.Description</td>
<td>description</td>
</tr>
<tr>
<td>Rights</td>
<td></td>
</tr>
<tr>
<td>Copyright</td>
<td>rights</td>
</tr>
<tr>
<td>Description</td>
<td>description</td>
</tr>
<tr>
<td>Relation</td>
<td></td>
</tr>
<tr>
<td>Kind</td>
<td>relation</td>
</tr>
<tr>
<td>Resource.Identifier.Catalog</td>
<td>isPartOf</td>
</tr>
<tr>
<td>Resource.Identifier.Entry</td>
<td>identifier</td>
</tr>
</tbody>
</table>

This mapping is a means to assure interoperability with many applications and educational institutions worldwide that adopt the DC metadata. Besides, DC is supported by default by some of the most well-known digital repository systems, like DSpace, EPrints and Digital Commons (Castagné, 2013).
The EMP ontology

In this section, we define the EMP binding, by specifying how a learning technology system will represent or use a metadata instance of an educational resource that follows the EMP. Ontologies have already been used to this end (see subsection Binding of Metadata Elements) and seem a very promising approach, given that they manage to transform the textual information captured by a metadata instance into a machine-understandable format.

Here, we describe the OWL binding of EMP, that is the ontological model used to represent the structure and characteristics of an LO, as these have been defined in section EMP: New Elements and Modifications. In order to build our EMP ontology we followed a widely-adopted methodology, proposed in Noy & McGuinness (2001). For its formal representation, we were based on OWL 2 whereas for creating and managing our ontology, we used the Protégé editor.

Ontology structure

The “LearningObject” class is a top class used to capture the notion of an LO, or an educational resource in general. The various characteristics of an educational resource are expressed as either classes or properties in the ontology. The elements Title, Language, Description, and Aggregation Level of the General category of the EMP, are expressed via the datatype properties “title,” “language,” “description” and “aggregation level” respectively. We opted for datatype- and not object- properties given that all of these elements simply assign values to some of the resource’s basic characteristics and express no correlations among other entities. More specifically, the “aggregation level” is an integer datatype property that can take values from 1 to 4. The “language” property can have as fillers any of the known language identifiers, like “en” for English or “el” for Greek.

The Contribute element of the Life Cycle category is represented by the “Contributor” class. Refinements of the Contribute element (Contribute:Date, Contribute:Entity:Affiliation, Contribute:Role) are expressed via the corresponding datatype properties (“contributeDate,” “affiliation” and “contributorRole,” respectively). “contributorRole” is an enumerated datatype property that can be filled with the values publisher, creator, and reviewer, all corresponding to several important roles in the life cycle of a learning resource. Instances of the class “Contributor” are related to instances of the “LearningObject” class via the “contributor” object property.

Figure 1. The class hierarchy in the LO ontology
To express the format of an educational resource (captured by the element Format of the Technical category), the “FileFormat” class is introduced in our ontological schema. In order to capture its top-four possible values (Text, Image, Streaming Media and Application) the corresponding subclasses were placed under the “FileFormat” class (see Figure 1). Their refinements (as presented in Table 2) became instances of these subclasses. The remaining elements of the Technical category – expressing the physical size, special software or hardware requirements, as well as the time duration for streaming media – are described by the datatype property “size,” “requirement” and “duration,” respectively.

The Learning Resource Type element, which is used to specify the different educational types of LOs, is captured by the “LearningResourceType” class. In the EMP, this element is associated with a predefined list of terms (see Table 3). Each such term is represented as an instance in the “LearningResourceType” class, with the exception of Self-Assessment, Activity, Simulation and Exercise, which have been placed as subclasses of the main “LearningResourceType” class. The ontological structure of the Learning Resource Type element is indicated in Figure 2.

![Diagram of Learning Resource Type](image)

**Figure 2.** The Learning Resource Type category as class in the EMP ontology

![Diagram of Object and Datatype Properties](image)

**Figure 3.** Object and Datatype Properties in the LO ontology
Self-Assessment, Activity, Simulation and Exercise are subclasses of the LearningResourceType class, whereas their refinements are captured as instances.

The elements of the EMP Education category, which express concepts like the intended end users, the instructional context, the age range of end users, as well as difficulty level and average learning time, are represented by object properties (“intendedEndUserRole,” “instructionalContext,” “age,” “difficulty,” “typicalLearningTime,” respectively – see Figure 3). Finally, the datatype property “rights” has been defined in order to express the copyright data that apply.

The potential relationships among LOs, as these have been described in section EMP: New Elements and Modifications, are represented as object properties; these properties can be used to correlate instances of the “LearningObject” class. All of them are sub-properties of the main property “isRelatedtoMA” (see Figure 3).

Combination with other ontologies

Some elements of the EMP have been captured as object properties in the EMP ontology, meaning that they correlate a resource with concepts (rather than literal values) originating from other supporting ontologies. Such supporting ontologies can be the LearningOutcome ontology presented in Kalou, Solomou, Pierrakeas & Kameas (2012), the well-known FOAF ontology (http://xmlns.com/foaf/spec/) and ontologies that represent knowledge domains.

In particular, the element Keyword of the EMP General category is expressed via the object property “subject.” This property links instances of the “LearningObject” class with class instances representing specific knowledge domain concepts. By this way, a network of interrelations among LOs and knowledge domains is created, thus allowing for the effective management and reusability of resources. Besides, it is crucial in a distance learning course to efficiently discover the appropriate educational material and provide it to the learners in the right order.

The “Contributor” class of our EMP ontology has been set as equivalent to the “Agent” class of the FOAF ontology. We, therefore, take advantage of the FOAF’s built-in object properties “foaf:name” and “foaf:surname” so as to formulate the Life Cycle element Contribute:Entity. In this way, the contributor of a resource is represented in an interoperable way and can be recognized in the context of other applications that also adopt the FOAF ontology.

Finally, in our EMP ontology we have introduced properties that combine LOs with concepts from the LearningOutcome ontology, to be used by LOs that have been designed to satisfy specific learning outcomes. More specifically, we determined the object properties “satisfies” and “satisfiesInd,” both of which associate instances of the “LearningObject” class with instances from “LearningOutcome” class, directly and indirectly, respectively. The indirect correlation of LOs can be only elaborated by restriction rules, expressed in the Semantic Web Rule Language (SWRL), and by exploiting existing reasoning mechanisms. By using both of these properties, we can better handle LOs within a distance learning course, given that we can retrieve them based on their association to the goals of learning.

Evaluation of the EMP ontology

To evaluate the potential of the proposed EMP, we exploited it for describing LOs designed to serve several distance learning courses of HOU. In particular, we have taken advantage of the EMP’s ontological binding and used it to characterize LOs for the following courses: Object Oriented Programming (Java), C Programming Language, Introduction to Computer Science, Programming Techniques and Data Structures.

In this section, we present an example use of the EMP ontology in the knowledge domain of Object Oriented Programming (Java). Our main goal is to illustrate how such an ontological binding of an educational metadata schema could lead to a better organization and discovery of resources within intelligent e-learning applications and allow for their reusability among different educational contexts.
Population of the EMP ontology

The Java programming language is a domain of knowledge covered by the HOU study course module of “Software Engineering.” An ontological representation for this knowledge domain is already available and is described in Kouneli, Solomou, Pierrakeas & Kameas (2012).

With the aid of the Protégé editor, we combined the EMP ontology with the Java ontology and the LearningOutcome ontology, presented in Kalou et al. (2012). The resulting (combined) ontology constitutes a rich knowledge representation, capable of effectively characterizing, organizing and correlating LOs and learning outcomes for the Java course, via semantic relationships.

The resulting combined ontology was populated with real individuals and in particular with LOs that have been developed for the course of Java. For every LO we created an instance in the combined ontology, by making the corresponding object- and datatype- assertions for it. We finally got 24 instances in total. An example LO instance, as this is captured in Protégé, is depicted in Figure 4.

Using the “subject” property we correlated a LO with concepts from the knowledge domain of the Java programming language (Figure 4). Nevertheless, in the place of Java, any other knowledge domain ontology could be imported, and thus render our combined ontological schema able to handle LOs of a different course.

![Figure 4. Example instance of LO with ID “LO_PLH_24_E2_17” in the combined ontology](image)

The property “satisfies” links an LO with instances of the LearningOutcome ontology. Consequently, the various semantic relationships that have been declared among the learning outcome instances can yield implicit – and very useful in terms of discovery – relationships for the corresponding LOs.
Example queries

To determine the capabilities of the proposed ontological model we ran semantic queries, and evaluated them against the populated ontology. The aim was to examine our model’s capability to infer knowledge. These queries are to be used in the context of an intelligent tutoring system, able to take advantage of this model, so they are expressed in a formal semantic query language. More specifically, they are expressed in the Manchester OWL Syntax (Horridge & Patel-Schneider, 2008) and for the time being they have been tested against the DL query tab of Protégé. We actually want to demonstrate that apart from running simple lookup queries, based on matching mere literal values, we can request more complex answers, based on the semantic relationships that an ontology allows.

Consider, for example, that a Learning Management System operating on behalf of a tutor or, as would be the most frequent case in Distance Learning, by a student, has to obtain all LOs that are difficult to learn, satisfy learning outcomes that concern Java operators and fall into the Cognitive Domain of the Bloom Taxonomy. This is expressed by the query#1 of Table 5. With query#2, we can retrieve those LOs having as learning resource type narrative text. Finally, with query#3, we can retrieve all LOs that either explicitly (through “satisfies” property) or implicitly (through “satisfiesInd” property) lead to the completion of learning outcomes related to Java operators. The results of these queries, as evaluated through the query tab of Protégé, can be seen in Figure 5.

![Figure 5. Results retrieved after evaluating queries 1, 2 & 3 of Table 5](image)

Table 5. Some Example Queries in Manchester OWL Syntax

<table>
<thead>
<tr>
<th>#</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>satisfies some (LearningOutcome and subject some Operator and domain value CognitiveDomain) and difficulty value “difficult”</td>
</tr>
<tr>
<td>2</td>
<td>learningResourceType value narrative_text</td>
</tr>
<tr>
<td>3</td>
<td>satisfies some (LearningOutcome and subject some Operator) and satisfiesInd some (LearningOutcome and subject some Operator)</td>
</tr>
</tbody>
</table>

Of course similar requests can be made for different domain subjects, different levels of knowledge and any kind of relationship modeled in the ontology as a property. These semantic queries are actually examples of competency
questions for the proposed ontology. Competency questions are a commonly used technique for evaluating ontology based formalisms (Grüninger & Fox, 1995).

Conclusions

In this work we propose an educational metadata profile (EMP) that can be used to characterize digital educational resources intended for use in distance learning courses. The proposed EMP either modifies some of the existing elements of the IEEE LOM schema, in order to reflect important aspects of the educational material, such as its technical data type or learning resource type, or augments it with new elements, such as the expected learning outcomes of a course, thus expressing concepts which are essential in the learning process, especially if it is supported by e-learning systems. EMP is rich enough to effectively describe both educational and technical aspects of an educational resource and especially a LO, but not exceedingly analytic so as to become difficult in use. We model our EMP as an ontology, in order to capture and process the semantic relationships among learning resources. Ontologies come with many applications in the field of education and their usage for representing the structure of an educational resource could lead to the development of advanced, intelligent applications. The ontological representation of our EMP was accomplished by “translating” its structural elements to classes, properties and instances in the ontology.

We applied EMP to characterize real LOs, designed for the HOU distance course on Object Oriented Programming (Java). With the aid of Protégé we show that the ontological representation of our EMP could lead to the discovery of LOs according to a specific concept of the knowledge domain – even if this is not explicitly stated - or according to very specific characteristics of a course, i.e., learning outcomes. This knowledge-based discovery could facilitate the process of designing a course and reusing LOs in various educational contexts, as described in Pierrakeas, Solomou and Kameas (2012).

The proposed EMP is already being used for the characterization of the HOU educational resources (including textbooks, digital supplementary material, presentations, theses, etc.). In addition, it is one of the core elements of the new LO-based approach adopted by HOU for the delivery of educational content to its students. Currently HOU is developing a Personalized Learning System that will use this EMP and its ontological representation in order to offer advanced services to distance learners. At the same time, HOU is investing in the development of knowledge domain and learning outcome ontologies for its courses; these will be gradually integrated to the EMP ontologies that are being developed for the corresponding LOs. Then, we plan to utilize an instructional design methodology to organize e-learning courses using LOs that are described by EMP, thus providing advanced services for learners that could support efficiently the handling and dissemination of educational material according to their specific needs.

References


From Motivation to Engagement: The Role of Effort Regulation of Virtual High School Students in Mathematics Courses

ChanMin Kim1, Seung Won Park2, Joe Cozart3 and Hyewon Lee1

1Learning, Design, and Technology, University of Georgia, Athens, Georgia, USA // 2Department of Medical Education, Sungkyunkwan University, South Korea // 3Georgia Virtual Learning, Georgia Department of Education, Atlanta, Georgia, USA // chanmin@uga.edu // parkswon@skku.edu // joe.cozart@gavirtualschool.org // ehyewon@uga.edu

*Corresponding author

(Submitted June 5, 2014; Revised December 10, 2014; Accepted May 3, 2015)

ABSTRACT

Engagement and motivation are not one and the same, but motivation can be transformed into engagement with proper design of support. In this study, we examined the differences between high performers and low performers with regard to changes in their motivation, regulation, and engagement throughout the semester. Participants were 100 students enrolled in online self-paced asynchronous mathematics courses offered at a virtual high school in the United States. A survey was administered to participants at three times throughout the semester. Data were analyzed using repeated measures MANOVAs. Overall, high performers and low performers differed with regard to their changes in motivation and regulation throughout the course, specifically, in self-efficacy and effort regulation. The study findings offer implications for teaching and research on creating potentially effective support for virtual learning.

Keywords

Virtual high school, Motivation, Regulation, Engagement, Mathematics education

Introduction

In recent years, online education has drastically increased, including at the K-12 level (Watson, Murin, Vashaw, Gemin, & Rapp, 2011). The enrollment of K-12 school students in online courses continues to grow along with the popularity of virtual schooling (Tucker, 2007). Every state in the United States and the District of Columbia has a K-12 virtual school (Kennedy & Archambaut, 2012). The rapid growth of virtual schooling has been attributed to numerous factors, especially its perceived benefits such as provision of individualized instruction and broadening educational access (Barbour & Reeves, 2009). The effectiveness of online and face-to-face education is now largely considered equal, which may have added momentum to the growth of virtual schooling (Hughes, McLeod, Brown, Maeda, & Choi, 2007). However, as with face-to-face schooling, high enrollment does not necessarily imply a high success rate. Challenges in virtual schooling include low performance and high course dropout rates (Barbour & Reeves, 2009).

Motivation is critical in learning. This is no less true in online learning (Carpenter & Cavanaugh, 2012). However, motivated students do not always engage in learning (Keller, 2008). Motivation to learn is only a desire to be involved in activities for learning (Kim & Bennekink, 2013). What makes students actually learn is their mindful engagement in those learning activities because “engagement leads to outcomes such as achievement” and “motivation underpins engagement” (Martin, 2012, p. 305).

There has been much research on motivation and engagement in a variety of face-to-face learning contexts (e.g., Junco, Elavsky, & Heiberger, 2013). However, what has been learned from such research may not apply to virtual schooling because of the unique characteristics of online learning environments (Cho, Demei, & Laffey, 2010) such as the lack of social presence, defined as “the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationship” (Short, Williams, & Christie, 1976, p. 65). Social presence and its lack have been researched in many studies to understand learning processes in online courses (e.g., Shea & Bidjerano, 2010). Student motivation can be different depending on the quantity and quality of social presence (Borup, Graham, & Davies, 2012; Shea & Bidjerano, 2010). This may apply even more to adolescents who tend to heavily weigh the importance of peers (Berten, 2008). In fact, the K-12 online education literature highlights the role of students’ interactions with their instructor and classmates (e.g., DiPietro, Ferdig, Black, & Preston, 2008).
addition, mathematics educators and researchers have underscored social aspects of mathematics learning (Davydov & Kerr, 1995).

In sum, there is a need to understand how students’ motivation and engagement influence their achievement in virtual high school mathematics courses so that support can be planned and implemented accordingly. The aims of the present study were to (a) explore and document how students’ motivation and engagement were related to their mathematics achievement at a virtual high school and (b) determine what support is needed in order to improve their motivation, engagement and achievement. This research can potentially provide a new lens through which to view how motivation and engagement interrelate with student achievement in virtual schooling. In the following sections, we discuss the definitions of engagement with an emphasis on its difference from motivation. We then discuss what is needed to transform motivation into engagement. Our research question is then posed.

What is engagement?

There is no straightforward way of defining the construct of engagement. Rather, it may be reasonable to define engagement as a multi-component construct comprised of subsets with associated indices. The engagement definition of Fredricks, Blumenfeld, and Paris (2004) encompasses three kinds of engagement: behavioral, cognitive, and emotional engagement. Behavioral engagement refers to involvement in learning tasks and environments such as time-on-task and attendance; cognitive engagement refers to psychological investment in the process of learning such as the use of learning strategies; and emotional engagement refers to affective reactions to learning tasks and environments such as emotions (Fredricks et al., 2004). The multi-component approach to considering engagement as a meta-construct can be conceptually and practically useful in research on and development of interventions to improve student engagement (Fredricks et al., 2004). Such an approach can broaden understanding of engagement (Finn & Zimmer, 2012; Fredricks et al., 2004; Lawson & Lawson, 2013). For example, if students’ emotional experience is examined along with their off-task behaviors such as disrupting a peer (Skinner, Kindermann, & Furrer, 2008), one could better understand how to improve their engagement by providing relevant support for negative emotions such as boredom.

In the present study, we define engagement as cognitive and affective participation in learning activities. We included only cognitive engagement (i.e., using shallow and deep cognitive strategies; Pintrich, Smith, Garcia, & McKeachie, 1993) and emotional engagement (i.e., experiencing boredom, anxiety, enjoyment, anger, shame, pride, and hopelessness; Pekrun, Goetz, & Frenzel, 2005) in our definition. We recognize that behavioral engagement is critical. However, in asynchronous online education, there are no face-to-face or synchronous virtual classes to attend and thus, the notion of behavioral engagement is not conceptually clear. For example, students’ login time does not necessarily mean how many hours they studied. They may log in just to download course materials. In addition, although Fredricks and her colleagues’ (2004) view of engagement as a meta-construct was applied in the present study, we excluded motivation from cognitive engagement unlike their definition of cognitive engagement. Engagement does not occur without desire to engage (Martin, 2012) but engagement and motivation are not one and the same.

How can motivation be transformed into engagement?

Motivation and engagement do not always coexist. In other words, there could be motivation but without engagement (e.g., only wanting something but not actually doing it). What transforms motivation to engagement is the effort and metacognitive regulation that students put into the process of their learning (Pintrich et al., 1993). Effort regulation is to control “one’s effort expenditure” (Halisch & Heckenhausen, 1977, p. 724). Metacognitive regulation is to control “one’s own cognition” (Pintrich et al., 1993, p. 803). Effort regulation is part of resource management (Pintrich et al., 1993). To display the role of the effort and metacognitive regulation in transforming motivation to engagement, here is an example. Reviewing class notes over and over (i.e., rehearsal, one of the cognitive strategies) is one way to engage in learning activities (Fredricks et al., 2004). This action of rehearsal (i.e., engagement) would not happen without the desire to learn (i.e., motivation); at the same time, the desire alone does not guarantee engagement and the student should also make an effort to rehearse (i.e., effort regulation) and monitor
when to rehearse (i.e., metacognitive regulation). Managing both cognition (i.e., metacognitive regulation) and effort (i.e., effort regulation) is important in learning (Pintrich et al., 1993) because it transforms motivation to engagement. Such regulation happens more easily when students engage in the learning tasks that are (a) perceived easy to execute and (b) interesting and enjoyable. Self-efficacy is defined as one’s perceived ability to successfully complete a task (Bandura, 1977). Intrinsic task value is defined as the value one perceives in a task that is inherently interesting and enjoyable (Schunk, Pintrich, & Meece, 2008). In many different learning environments, self-efficacy has been steadily found to be a strong predictor for motivation and performance (e.g., Multon, Brown, & Lent, 1991). Self-efficacious students also tend to control their learning process (Bandura, 1977). Thus, when a task is perceived to be easy to perform, students are likely to perceive high self-efficacy and to self-regulate. Self-efficacy influences motivation directly and engagement indirectly (Schunk & Mullen, 2012). Students engage in tasks also for their own interests (Ainley, 2012) and enjoyment (Csikszentmihalyi, 1988) when the intrinsic value of the tasks is high (e.g., Deci & Ryan, 2008).

Not every student enjoys mathematics. Still, students can engage in learning tasks for which they do not perceive high intrinsic value when there is no obstacle that they believe they cannot overcome. In other words, when students have high expectancy of success (Wigfield & Eccles, 2000), motivation can be transformed into engagement. However, not every task is easy. Especially in online mathematics courses, not only do many students not enjoy math, but also they are not self-efficacious due to previous failure of math courses. Thus, such students often experience negative emotions like anger in math classes (Kim, Park, & Cozart, 2014).

Research question

This study investigated how differently virtual high school students engage and achieve in mathematics courses and what quality of theirs makes such differences. We addressed the following research question: How do high performers and low performers differ with regard to their changes in motivation, regulation, and engagement throughout the course? We compared such changes from the beginning of the semester to the middle of and the end of the semester. In this study, motivation variables included self-efficacy and intrinsic value, regulation variables included metacognitive regulation and effort regulation, and engagement variables included cognitive engagement (i.e., using deep cognitive strategy use and shallow cognitive strategy use) and emotional engagement (i.e., experiencing boredom, anxiety, enjoyment, anger, shame, pride, and hopelessness). Table 1 summarizes each construct and variable.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Construct definition</th>
<th>Variable</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Desire to be involved with learning activities/tasks</td>
<td>Self-efficacy</td>
<td>Beliefs about own abilities to complete learning tasks in a certain circumstance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intrinsic value</td>
<td>Perception of the value of learning tasks in relation to his or her interest</td>
</tr>
<tr>
<td>Regulation</td>
<td>Management of cognition and other resources such as effort, emotions, and environments</td>
<td>Metacognitive regulation</td>
<td>Management of cognition in learning activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Effort regulation</td>
<td>Management of effort in learning activities in the face of difficulties</td>
</tr>
<tr>
<td>Engagement</td>
<td>Cognitive and affective participation in learning activities</td>
<td>Cognitive engagement</td>
<td>Involvement with learning activities using shallow and deep cognitive strategies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emotional engagement</td>
<td>Emotional reactions, such as boredom, anxiety, enjoyment, anger, shame, pride, and hopelessness, to learning activities</td>
</tr>
</tbody>
</table>

Table 1. Variable description (Operationalization of the constructs in this study)
Methods

Participants and setting

Participants were students enrolled in online self-paced asynchronous mathematics courses offered at a virtual high school in the southeastern United States. The virtual high school is run by the State Department of Education. Students who are enrolled in the virtual high school courses either take courses for an entire curriculum or supplement courses that they take at their local school. A survey was administered to participants at three times throughout the semester. One hundred participants who completed the survey all three times were included in the study. The participants \((n = 100)\) were from Math 1 \((n = 13)\), Math 2 \((n = 4)\), Math 3 \((n = 9)\), Algebra \((n = 31)\), Geometry \((n = 7)\), Pre-Calculus \((n = 5)\), Calculus \((n = 14)\), Statistics \((n = 16)\), and Applied Math (i.e., Problem Solving and Money Management) \((n = 1)\) courses. The average age was 15.9. Sixty-eight out of 100 were female. 71% of the participants were Caucasian, 11% were Black/African American, 8% were Asian American, 3% were Hispanic/Latino, and 7% were multiracial. Those who had no prior experience with online math courses \((n = 82)\) outnumbered those with experience.

Data collection

The Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, & DeGroot, 1990) was used to measure motivation, regulation, and cognitive engagement. Participants responded to each of the 40 items using a 7-point Likert scale ranging from (1) “Not at all true of me” to (7) “Very true of me.” The reliability of scores on these subscales of the MSLQ ranged from .52 to .93 (Pintrich, Smith, Garcia, & McKeachie, 1991) and validity of the items were tested in a variety of school settings (e.g., Wolters & Pintrich, 1998). Some items were reworded to reflect the online context of this study (e.g., the item “When reading I try to connect the things I am reading about with what I already know” was revised to “when reviewing online course materials I try to connect the things I am reviewing with what I already know”). Scale reliability coefficients with the reworded items ranged from .59 to .90 in a previous study (Kim et al., 2014) and from .50 to .88 in the current study (see Table 2).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Scale</th>
<th>Sample item</th>
<th>Scale reliability (Cronbach’s α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Self-efficacy</td>
<td>I am sure I can do an excellent job on the problems and tasks assigned for this class.</td>
<td>.88</td>
</tr>
<tr>
<td></td>
<td>Intrinsic value</td>
<td>Even when I do poorly on a test I try to learn from my mistakes.</td>
<td>.88</td>
</tr>
<tr>
<td>Regulation</td>
<td>Metacognitive regulation</td>
<td>Before I begin studying I think about the things I will need to do to learn.</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>Effort regulation</td>
<td>I work hard to get a good grade even when I don’t like a class.</td>
<td>.66</td>
</tr>
<tr>
<td>Engagement</td>
<td>Deep cognitive strategy use</td>
<td>When reviewing online course materials I try to connect the things I am reviewing with what I already know.</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>Shallow cognitive strategy use</td>
<td>When I study for a test I try to remember as many facts as I can.</td>
<td>.73</td>
</tr>
<tr>
<td></td>
<td>Boredom</td>
<td>Just thinking of my math homework assignments makes me feel bored.</td>
<td>.88</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>I’m so scared of my math assignments that I would rather not start them.</td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td>Enjoyment</td>
<td>The material we deal with in mathematics is so exciting that I really enjoy my class.</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>I am so angry that I would like to throw my homework into the trash.</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>Shame</td>
<td>I feel ashamed when I realize that I lack ability.</td>
<td>.83</td>
</tr>
</tbody>
</table>
After having done my math homework, I am proud of myself. 

I would prefer to give up.

The Achievement Emotion Questionnaire in Mathematics (AEQ-M) (Pekrun, Goetz, & Frenzel, 2005) was used to measure emotional engagement. Nineteen items were excluded from the current study because they were pertinent to attending a physical classroom (e.g., “When I say something in my math class, I can tell that my face gets red.”). Only the items asking about emotional experiences before, during, and after studying (18 items) and taking an exam (23 items) were included. Participants responded to each of 41 items using a 5-point Likert scale ranging from (1) “Strongly disagree” to (5) “Strongly agree.” Some items were reworded to reflect the online context of this study. Internal consistency coefficients of scores on the various sub-scales of the AEQ-M ranged from .84 to .92 in a previous study (Frenzel, Pekrun, & Goetz, 2007) and validity of the scores was tested in a variety of applications (e.g., Frenzel, Thrash, Pekrun, & Goetz, 2007). Scale reliability coefficients with the reworded items ranged from .67 to .93 in a previous study (Kim et al., 2014) and from .75 to .93 in the current study (see Table 2).

Achievement was measured using students’ final grades. The possible range of the final grades was 0 – 100. Final grades were determined using scores from asynchronous discussions, homework assignments, quizzes, tests, and the final exam. There was no grade directly tied to attendance. Each course used a standard weighting system to distribute grades across discussions, assignments, quizzes, tests, and exams.

Procedure

We recruited participants in the first and second weeks of the Fall 2011 semester. In the course website, we posted a URL of a webpage containing an online survey that includes (a) the study description, (b) consent forms, (c) demographic questions, and (d) 1st survey questions on motivation, regulation, and engagement. Students who submitted signed parental consent and student assent forms proceeded to respond to demographic and 1st survey questions. The same survey on motivation, regulation, and engagement was administered two more times throughout the semester: one was in the middle of the semester and the other was toward the end of the semester. We collected the final grade scores of the participants when the semester ended. Figure 1 illustrates the two groups, three measurement points, and six variables of the study.

![Figure 1. A summary of data collection](image)

Data analyses

Four separate 3 (time) × 2 (group) MANOVAs were conducted with time (Measurement Points 1, 2, and 3) as a repeated measure to investigate differences in changes in motivation, regulation, cognitive engagement, and
emotional engagement between the high-performer and low-performer groups. The participants were categorized into high, middle, and low performer groups based on their final grade scores ($M = 79.11, SD = 19.43$). Because the variability of the final scores was relatively high, we were concerned that grouping the participants based upon the mean ± one standard deviation may not include students who scored high enough to be considered as high performers. Thus, we categorized participants using the conventional letter grade assignment: participants with final grade scores higher than 90 (equivalent to a letter grade A) were included in the high-performer group ($M = 94.13, n = 38$) while participants with final grade scores lower than 80 (or, those who received a letter grade C or below) were included in the low-performer group ($M = 61.68, n = 40$). The rest were regarded as the middle performers. For the purpose of examining differences between high performers and low performers, the middle performers were excluded in these analyses. Partial eta-squared ($\eta^2_p$) was used to calculate effect size: Small: $0.01 \leq \eta^2_p < 0.06$; Medium: $0.06 \leq \eta^2_p < 0.14$; Large: $\eta^2_p \geq 0.14$.

Results

The results of repeated measures MANOVAs indicated that high performers and low performers differed with regard to their changes in motivation, regulation, and engagement throughout the course, specifically, in self-efficacy (part of motivation) and effort regulation (part of regulation). The descriptive statistics of all dependent variables examined are presented in Table 3. The analysis results of the repeated measures are summarized in Table 4.

<table>
<thead>
<tr>
<th>Measurement time point</th>
<th>Low performer ($n = 40$)</th>
<th>High performer ($n = 38$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Self-efficacy(^a)</td>
<td>41.40 (8.09)</td>
<td>37.70 (10.57)</td>
</tr>
<tr>
<td>Intrinsic value(^b)</td>
<td>46.90 (8.14)</td>
<td>43.12 (9.65)</td>
</tr>
<tr>
<td>Effort regulation(^c)</td>
<td>19.00 (3.63)</td>
<td>18.62 (4.18)</td>
</tr>
<tr>
<td>Meta. regulation(^d)</td>
<td>22.97 (3.87)</td>
<td>22.30 (4.28)</td>
</tr>
<tr>
<td>Deep strategy(^e)</td>
<td>40.57 (5.57)</td>
<td>40.90 (6.61)</td>
</tr>
<tr>
<td>Shallow strategy(^f)</td>
<td>25.12 (5.07)</td>
<td>25.10 (6.09)</td>
</tr>
<tr>
<td>Boredom(^g)</td>
<td>8.72 (3.94)</td>
<td>9.27 (4.00)</td>
</tr>
<tr>
<td>Anxiety(^h)</td>
<td>35.80 (11.44)</td>
<td>34.40 (8.41)</td>
</tr>
<tr>
<td>Enjoyment(^i)</td>
<td>17.05 (4.66)</td>
<td>17.20 (3.22)</td>
</tr>
<tr>
<td>Anger(^j)</td>
<td>12.92 (5.06)</td>
<td>13.45 (4.60)</td>
</tr>
<tr>
<td>Shame(^k)</td>
<td>14.00 (5.30)</td>
<td>12.82 (4.14)</td>
</tr>
<tr>
<td>Pride(^l)</td>
<td>13.42 (3.72)</td>
<td>13.65 (2.96)</td>
</tr>
<tr>
<td>Hopelessness(^m)</td>
<td>18.45 (7.05)</td>
<td>17.12 (6.96)</td>
</tr>
</tbody>
</table>

Notes. \(^{a}\)Possible range of Self-Efficacy score: 9-63; \(^{b}\)Possible range of Intrinsic Value score: 9-63; \(^{c}\)Possible range of Effort regulation: 4-28; \(^{d}\)Possible range of Metcognitive Regulation: 5-35; \(^{e}\)Possible range of Deep Cognitive...
Strategy: 8-56; Possible range of Shallow Cognitive Strategy: 5-35; Possible range of Boredom score: 3-15; Possible range of Anxiety score: 11-55; Possible range of Enjoyment score: 6-30; Possible range of Anger score: 5-25; Possible range of Shame score: 5-25; Possible range of Pride score: 4-20; Possible range of Hopelessness: 6-30.

Table 4. Summary of univariate analyses of repeated measures

<table>
<thead>
<tr>
<th></th>
<th>Group effect</th>
<th>Time effect</th>
<th>Time x Group effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>22.74</td>
<td>.000</td>
<td>10.07</td>
</tr>
<tr>
<td>Intrinsic value</td>
<td>1.89</td>
<td>.173</td>
<td>26.57</td>
</tr>
<tr>
<td>Effort regulation</td>
<td>12.52</td>
<td>.001</td>
<td>4.91</td>
</tr>
<tr>
<td>Meta. regulation</td>
<td>.192</td>
<td>.662</td>
<td>3.45</td>
</tr>
<tr>
<td>Deep strategy</td>
<td>N/A</td>
<td>N/A</td>
<td>5.11</td>
</tr>
<tr>
<td>Shallow strategy</td>
<td>N/A</td>
<td>N/A</td>
<td>6.71</td>
</tr>
<tr>
<td>Boredom</td>
<td>N/A</td>
<td>N/A</td>
<td>1.610</td>
</tr>
<tr>
<td>Anxiety</td>
<td>N/A</td>
<td>N/A</td>
<td>.011</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>N/A</td>
<td>N/A</td>
<td>2.66</td>
</tr>
<tr>
<td>Anger</td>
<td>N/A</td>
<td>N/A</td>
<td>1.00</td>
</tr>
<tr>
<td>Shame</td>
<td>N/A</td>
<td>N/A</td>
<td>1.34</td>
</tr>
<tr>
<td>Pride</td>
<td>N/A</td>
<td>N/A</td>
<td>5.27</td>
</tr>
<tr>
<td>Hopelessness</td>
<td>N/A</td>
<td>N/A</td>
<td>3.03</td>
</tr>
</tbody>
</table>

Note. Significant effects are in bold.

The first 3 (time) × 2 (group) repeated measures MANOVA was conducted with two motivation variables: self-efficacy and intrinsic value. One important assumption of a repeated measures MANOVA is the equality of covariance. Results of Box’s Test of Equality Covariance Matrices yielded $X^2(21) = 36.76, p = .01$, providing evidence of a violation of the equal covariance assumption. Nevertheless, because the natural logs of covariance matrices were found to be similar, we proceed with the usual MANOVA tests following Huberty and Olejnik’s (2006) suggestion. Preliminary analyses were conducted to examine if there were any differences between two groups at the beginning of the semester (e.g., Time 1), and we found a significant difference in self-efficacy ($p < .01$) between two groups: high performers demonstrated higher self-efficacy than low performers at Time 1.

On the main analysis of the repeated measures MANOVA with two motivation variables, there were a significant main effect of time, Wilks’ Lambda = .579, $F(4, 73) = 13.23, p < .001$, a main effect of group, Wilks’ Lambda = .759, $F(2, 75) = 11.86, p < .001$, and a significant time x group interaction, Wilks’ Lambda = .847, $F(4, 73) = 3.29, p < .05$. To further inspect the significant effects on the multivariate analysis, follow-up univariate analyses of repeated measures were conducted for each motivation variable. Follow-up univariate analyses for self-efficacy yielded a significant main effect of time, $F(2, 75) = 10.07, p < .001$, $\eta^2 = .21$, a main effect of group, $F(1, 76) = 22.74, p < .001$, $\eta^2 = .23$, and a significant time x group interaction, $F(2, 75) = 6.32, p < .01$, $\eta^2 = .14$. Further analyses indicated that self-efficacy among the low-performer group gradually diminished from Time 1 to Time 3 ($p < .001$); the self-efficacy of the high-performer group did not change over time. Last, follow-up univariate analyses for intrinsic value yielded a significant main effect of time, $F(2, 75) = 26.57, p < .001$, $\eta^2 = .41$, indicating that both high- and low-performer groups reported a gradual decrease in intrinsic value over three measurement times.

The second 3 (time) × 2 (group) repeated measures MANOVA was conducted with two regulation variables: metacognitive regulation and effort regulation. The equality of covariance matrices was upheld as indicated by $X^2(421) = 30.65, p = .07$. Preliminary analyses indicated a significant difference in effort regulation between two groups ($p < .05$) at Time 1: high performers showed significantly higher effort regulation at Time 1 than low performers. Results of the repeated measures MANOVA revealed a significant main effect of time, Wilks’ Lambda = .851, $F(4, 73) = 3.17, p < .05$, and a main effect of group, Wilks’ Lambda = .757, $F(2, 75) = 12.02, p < .001$. Follow-up univariate analyses for effort regulation yielded a significant main effect of time, $F(2, 75) = 4.91, p < .05$, $\eta^2 = .11$, and a main effect of group, $F(1, 76) = 12.52, p < .01$. While high performers maintained superior effort regulation to low performers throughout the semester, both groups demonstrated diminished effort regulation from Time 1 to Time 3 ($p < .01$, $\eta^2 = .11$). Similarly, univariate analyses for metacognitive regulation indicated that both high and low performers gradually reported lesser metacognitive regulation from Time 1 to Time 3 ($p < .05$, $\eta^2 = .08$).
The third 3 (time) × 2 (group) repeated measures MANOVA was conducted with two cognitive engagement variables: deep strategy use and shallow strategy use. Preliminary analyses indicated that high and low performers demonstrated the similar level of both deep and shallow strategy use at Time 1. Given the equal covariance indicated by \( \chi^2(21) = 26.44, p = .18 \), a significant main effect of time was found from the repeated measures MANOVA, Wilks’ Lambda = .829, \( F(4, 73) = 3.75, p < .01 \). Follow-up univariate analyses for deep and shallow strategies also yielded a significant time effect \( (p < .01, \eta^2 = .12; p < .01, \eta^2 = .15 \); respectively) indicating that both high and low performers decreased their use of deep and shallow strategies over time.

The last 3 (time) × 2 (group) repeated measures MANOVA analysis was conducted with seven emotional engagement variables: boredom, anxiety, enjoyment, anger, shame, pride, and hopelessness. Results of Box’s Test of Equality Covariance Matrices provided the evidence of covariance equality. Preliminary analyses indicated that the high and low performers differed in the level of shame \( (p < .01) \), pride \( (p < .05) \), and hopelessness \( (p < .01) \) in the beginning of the semester. Two groups were not different in the levels of any other emotion variables at Time 1. Results of the repeated measures MANOVA with emotional engagement variables indicated a significant time effect, Wilks’ Lambda = .669, \( F(14, 63) = 2.23, p < .05 \). Conducting follow-up univariate analyses, pride was the only variable that yielded the main effect of time, \( F(2, 75) = 5.27, p < .01, \eta^2 = .12 \). Both high and low performers diminished pride over time.

**Discussion**

**Findings and interpretations**

First, we found that high performers and low performers differed throughout the course: high performers started the semester with the higher level of effort regulation than low performers and they maintained their superior level of effort regulation to low performers’ throughout the semester. The higher the level of effort regulation that students had, the higher their achievement was. This finding is aligned with that of Puzziferro’s study (2008) with community college students enrolled in liberal arts online courses. Even when students’ perception of intrinsic task value was low in the current study, those who reported greater effort regulation tended to perform better than those who reported the lower level of effort regulation.

Given these findings, supporting students’ effort regulation may be one way to help them do better in an online learning environment. Designing support for effort regulation could involve online instructors’ scaffolding for student effort regulation that includes monitoring and guiding student efforts (Cho & Shen, 2013). Volition theories and models can be helpful in creating such support as well since they explain how efforts can be better regulated (e.g., implementation intentions in Gollwitzer & Sheeran, 2006; action control in Kuhl, 1985). Volition refers to “a dynamic system of psychological control processes that protect concentration and directed effort in the face of personal and/or environmental distractions” (Corno, 1993, p.14).

Second, we found that the metacognitive regulation of both high performers and low performers decreased throughout the semester. This contradicts previous findings on the role of metacognitive regulation in online learning (Artino, 2007; Cho & Shen, 2013). However, this finding along with discussions on effort regulation above suggests that students’ effort regulation may have compensated for the impact of decreased metacognitive regulation on achievement. This supports the notion that achievement depends not only on “cognitive control and regulation, especially the different cognitive, metacognitive, and learning strategies that students may use to control their own cognition and learning” but also on “how students control their own motivation, emotions, behavior (including choice, effort, and persistence), and their environment” (Pintrich, 1999, p. 336). Although in the current study we did not examine students’ regulation of other aspects such as emotions and environment, the inclusion of effort regulation is an attempt to understand the path from student motivation to achievement. This attempt may be critical especially in online learning environments where more qualities are expected than just knowing how to study (e.g., cognitive strategy use) (Kim & Bennekin, 2013).

Third, high performers started the semester with higher self-efficacy than low performers. Low performers’ self-efficacy gradually diminished over time while there was no change in self-efficacy among high performers. The indirect effect of self-efficacy on achievement has been well documented (e.g., Multon, Brown, & Lent, 1991) also in the literature involving online learning contexts (e.g., Cho & Shen, 2013). It is conceivable that effort regulation
may have influenced self-efficacy (Komarraju & Nadler, 2013). The role of effort regulation as a mediator is pointed out in some studies (Artino, 2007; Cho & Shen, 2013; Shea & Bidjerano, 2010). Thus, combined with the finding on effort regulation, it seems that there could be other ways to promote self-efficacy than structuring learning environments to provide vicarious experiences, autonomy, clear expectations, goal specificity, and balanced task difficulty (Bandura, 1997; Jang, Reeve, & Deci, 2010; Locke & Latham, 2002). The effect of self-efficacy can be improved through effort regulation.

Fourth, there was no difference between high performers and low performers in intrinsic value. This finding is counter-intuitive and we can only speculate what has happened based on relevant literature. The motivation literature describes that people tend to be persistent when they perceive intrinsic value in a certain task that satisfies their interest (Ainley, 2012) and which they enjoy (Csikszentmihalyi, 1988). Such perceived intrinsic value enhances the quality of motivation (Deci & Ryan, 2008) and provides momentum for participating in the task. With enjoyment, full engagement can occur without even a conscious effort (e.g., flow experience; Csikszentmihalyi, 1988). Nonetheless, without enjoyment and interest in a given task, people can be still engaged in a task and come to a successful completion depending on regulatory styles (Ryan & Deci, 2000). Along this line of literature, our finding on intrinsic value may have arisen as such: (a) the learning environment may have allowed students’ perceived intrinsic value to fade away considering that both high and low performer groups showed gradual decreases in intrinsic value throughout the semester and (b) even without enjoyment and genuine interest, students with effort regulation could succeed considering high performers had superior effort regulation to low performers. Not every student has the capability to “reshape tasks and to make them more palatable” in suboptimal learning contexts (Corno & Kanfer, 1993, p. 302) and those without such a capability such as low performers in this study can be educated about how to optimize contexts for themselves (e.g., exercising effort regulation) (Byman & Kansanen, 2008).

Last, both high and low performers’ pride and uses of deep and shallow strategies significantly diminished throughout the semester. The use of shallow cognitive strategies should be better than nonuse but when shallow cognitive strategies are used without deep cognitive strategies, learning tends to stay at a shallow level. Waning pride may have been due to the decreased use of cognitive strategies and/or the lack of intrinsic value. However, resiliency occurs when negative emotions serve as a warning for students with clear goals (Turner, & Schallert, 2001). The steadily superior level of effort regulation that the high performers had may have allowed them to be resilient from decreased pride and still be successful in the course.

Implications for research and practice

The findings offer implications for research on and teaching at virtual schools. Understanding how students’ motivation and engagement as well as regulation contribute to their learning provides information of how support can be planned accordingly in virtual high school math courses. That is, comparing high and low performers’ changes in their motivation, regulation, and engagement provides direction for creating potentially effective support, especially for student effort regulation, for online education in K-12 virtual schools. For example, support for students’ effort regulation may help not only with a lack of motivation from not viewing the intrinsic value of learning tasks but also with disengagement such as nonuse of cognitive strategies, which would in turn improve achievement. This is a unique way of improving motivation, engagement, and achievement especially when every learning environment cannot be optimal for every student (Kinshuk, Liu, & Graf, 2009). Improving learning through effort regulation can also contribute to greater capacity for lifelong learning. Along this line, other qualities of students could be studied also. For example, students’ beliefs about intelligence “valuing effort and hard work” could be used to improve effort regulation (Komarraju & Nadler, 2013, p. 70). Also since self-efficacious students tend to believe that their performance can be improved by exerting effort (Komarraju & Nadler, 2013), improving self-efficacy can lead to improved effort regulation. Even when tasks are difficult, self-efficacious students tend to be persistent (Komarraju & Nadler, 2013).

Limitations and suggestions for future research

There are several limitations in this study. First, mainly self-reported data were used. The social desirability issue (Crowne & Marlowe, 1960) remains. Future research should consider individual or focus group interviews as well as
online behavioral observations using learning analytics and asynchronous communications such as emails. Second, differences among courses in which participants were enrolled were not investigated due to the small sample size per course. The study findings should be interpreted with caution especially due to these limitations that also make it hard to generalize the study findings to other US virtual school contexts. A study with a larger sample size would increase statistical power. Alternative sampling methods should be considered in future studies. Third, individual differences among participants may have contributed to the difference in performance such as prior knowledge, parental support, tutoring help, gender and socioeconomic status. Fourth, regulation of other resources such as motivation, emotions, and environment (Pintrich, 1999) was not investigated in the current study. Last, social presence was not empirically examined in this study to see if social presence actually lacks in the virtual learning environment of this study.

References


Automatic Scaffolding and Measurement of Concept Mapping for EFL Students to Write Summaries

Yu-Fen Yang
Graduate school of applied foreign languages, National Yunlin University of Science & Technology, Taiwan // yangy@yuntech.edu.tw

(Submitted June 26, 2014; Revised February 26, 2015; Accepted March 12, 2015)

ABSTRACT
An incorrect concept map may obstruct a student’s comprehension when writing summaries if they are unable to grasp key concepts when reading texts. The purpose of this study was to investigate the effects of automatic scaffolding and measurement of three-layer concept maps on improving university students’ writing summaries. The automatic three-layer concept maps presented in this study include: (1) the central idea of the title, (2) the main idea of each paragraph, and (3) the supporting ideas in each paragraph. A sample of 107 university students who studied English as a Foreign Language (EFL) was divided into experimental and control groups, comprised of 54 and 53 students respectively. The results of this study indicate that the students of the experimental group made more significant improvement in reading comprehension and summary writing than those of the control group as they were able to identify the main idea from each paragraph and clarify relations among paragraphs after requesting the three-layer concept maps. They then modified their prior knowledge structure by revising the first-draft summaries to final ones. The students’ perceptions toward the automatic three-layer concept maps to improve summary writing are also described in this study.

Keywords
Automatic scaffolding, Automatic measurement, Immediate feedback, Summary writing, Three-layer concept map

Introduction
A recent trend in evaluating students’ language proficiency has been the assessment of their integrated reading and writing skills rather than corresponding discrete skills (Kırmızı, 2009). University students who study English as a Foreign Language (EFL) are required to be equipped with both reading and writing skills in order to assist themselves in grasping main ideas or selecting key elements from reading a large quantity of texts in highly demanding academic courses (Dreyer & Nel, 2003). In acquiring reading and writing skills, students need to identify keywords or phrases from each paragraph, construct concept maps to decode texts, and integrate main ideas or key elements into organized and logical summaries (Kırmızı, 2009). Reading and writing success depends on word identification, cognition, construction, and summarizing skills (Sung, Chang, & Huang, 2008).

Writing summaries is a particularly difficult task for EFL students to learn as they have to determine what content in a passage is the most important and then transform it into succinct statements in their own words (Yang, 2014). In determining what information is important in texts, the main idea of a passage is often not present in the surface structure (the exact wordings) of the text (Friend, 2001; Kintsch, 1998) and the cognitive process which converts surface structure to an understanding of a text is internal and largely unobservable in onsite instruction (Alfassi, 2004; Fischer, 2003). As such, EFL novice summary writers may encounter difficulties as incorporating source text information into their own writing in terms of low reading comprehension skills (Esmaeili, 2002; Plakans, 2009) and the restriction of their vocabulary size (Baba, 2009).

To grasp the main ideas necessary for writing summaries, concept mapping has been reported to be an effective strategy which graphically indicates the relationships between multiple concepts (Tseng, Chang, Lou, Tan & Chiu, 2012). Liu, Chen and Chang (2010) have proposed that concept mapping strategies provide students with a more systematic and organized way to clarify the important concepts necessary for enhancing reading comprehension, thereby improving summarizing skills. In particular, concept mapping facilitates students externalize their prior knowledge and combine it with new ones for reconstructing new language knowledge and learning experiences (Novak, 1990; Hwang, Hung, Chen, & Liu, 2014).

In light of the rapid growth of computer technology, computerized concept maps are helping to better students’ reading comprehension and writing proficiency (Wu, Hwang, Milrad, Ke, & Huang, 2012). A concept map is a fill-
in-the-blank strategy, where students extract important words and phrases while reading a passage and then fill in the essential words or phrases in a map for comprehension and summary writing. Keyphrases, which include two or more keywords, are formed to represent important concepts (Mangina & Kilbride, 2008). Zha (2002) proposes that keyphrases and summaries comprise word-to-sentence relationships in terms of information retrieval generated by deleting unnecessary information, extracting keyphrases in each paragraph, and integrating keyphrases and main ideas to complete a summary. In forming concept maps, nodes refer to the main ideas, and links represent the associations between the main ideas (Adesope & Nesbit, 2013; Novak, 1993).

A concept map is also a useful assessment tool for measuring students’ reading comprehension and summary writing. Traditional paper-based concept maps cannot help the teacher immediately evaluate a student’s comprehension. As a result, students are unable to receive timely feedback from their teacher or peers. With the assistance of technology, computerized concept maps facilitate the modification of nodes and links and make the task easier for students to fill in the keyphrases, revise their previous maps, and automatically receive feedback for scoring (Liu, Chen, & Chang, 2010). As such, computerized concept maps not only provide a scaffold to help students visualize and grasp main ideas, and clarify any connected relations between them (Liu, 2011; Sturm & Rankin-Erickson, 2002), but they also offer students immediate feedback and scoring to modify their summaries without waiting a long time for teacher feedback (Wu, Hwang, Milrad, Ke, & Huang, 2012).

Self-regulated learning, which comes from cognitive psychology, refers to students consciously make efforts to manage and direct complicated learning activities in which they are required to learn independently (Zimmerman & Martinez-Pons, 1988). Self-regulated learning is defined as “an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features of the environment” (Pintrich, 2000, p. 453). By receiving immediate feedback and scoring from technology-enhanced language learning, students are provided with the opportunity to perform self-regulated learning without the restrictions of time and space (Wang, 2008). To foster students’ self-regulated learning, the automatic scaffolds of three-layer concept maps become effective tools to enable students to develop their language understanding beyond the given information, phasing out the teacher’s assistance as students’ reading and writing competence increases (Azevedo & Hadwin, 2005).

**Statements of problems**

While previous studies have examined the scaffolds of concept maps on reading and writing, few studies have focused on an explicit instruction which helps students construct their own maps when learning to write summaries. First, many studies (e.g., Cordero-Ponce, 2000; Friend, 2000; Irwin, 1991) have merely proposed general and unclear rules for writing a summary. Without many opportunities to observe their teacher or peers’ modeling on how to write summaries through log files, novice students are still unaware of what main ideas are missing or what irrelevant details are included in a summary (Liu, 2011).

Second, most studies generated concept maps only from corpora (e.g., Cimiano, Hotho & Staab, 2005). This makes the measurement of concept maps questionable since the various sizes of corpora may influence the complexities of concept maps. The criterion for evaluating students’ concept maps is not clearly revealed in previous studies as “each individual might have very different ways in generating his own concepts.” (Yang, Wong, & Yeh, 2008, p. 174).

Finally, many concept maps are developed using numerous key concepts with a large quantity of nodes and links. For example, Liu (2011) has proposed that using the concept mapping software program named Inspiration could improve ESL learners’ writing proficiency. However, complicated concept mapping in Inspiration did not provide students with the main idea in each paragraph. EFL students may still miss the key concept in each paragraph which hinders their comprehension when writing summaries as the software only presented the main concept of a text.

**Background of the study**

This study attempts to investigate the effects of automatic keyphrases, three-layer concept maps, and scoring on EFL students’ learning to write summaries. According to Novak and Gowin’s study (1984), a concept map consists of main concepts, sub-concepts, and important details from the central to the external within three layers. Many concept
mapping software programs provide students with the main concept to comprehend a reading passage. However, students still have difficulties writing summaries when using these software programs without a hierarchical structure for grasping the main idea in each paragraph. When the teacher’s support is decreased, immediate feedback, such as automatic map and scoring, becomes important for the students to clarify their text misunderstanding and evaluate the quality of maps for summary improvement.

This study developed an automatic three-layer concept map as scaffolding and measurement for university students’ learning to write summaries. Before forming a map, the system automatically extracted important keyphrases by self-developed algorithms. Then, according to the keyphrases, the three-layer concept maps was represented in three layers (see Figure 1) including (1) the central idea in the title, (2) the main idea in each paragraph, and (3) the supporting ideas in each paragraph. The numbers of nodes (main idea) in the second layer depend on how many paragraphs are included in a text. Students can write their first sentence of a summary based on the keyphrases in the title, and then complete the following sentences in terms of the main idea from each paragraph. As students completed their maps, the system automatically measured their maps by providing immediate feedback of scoring. The criterion of scoring is to compare the automatic map generated from the system.

To fulfill the research purpose of this study, which was to investigate the effects of automatic keyphrases, three-layer concept maps, and scoring on university students’ learning to write summaries, three research questions were addressed: (1) To what extent does students’ reading progress differ between a group using automatic three-layer concept maps and a group using paper-based concept maps? (2) To what extent does the students’ use of automatic three-layer concept maps help students learn to write summaries and enhance their reading comprehension? (3) What are the students’ perceptions of the value of using automatic three-layer and paper-based concept maps for improving their reading comprehension and summary writing?

![Figure 1. New three-layer concept map in this study](image)

**Method**

**Participants**

A total of 107 first year university students at a technological university in central Taiwan were recruited to learn how to write summaries and had their work compared to determine whether there was any improvement for writing summaries and reading comprehension after using automatic scaffolds and measurement of three-layer concept maps and paper-based concept maps. They undertook the course “Reading and Writing” to improve their academic proficiency. According to the results of the university survey, most of them (94%) did not have any experience in writing summaries in English or had attended online summary instruction. The students were set biweekly
assignments to read multiple types of academic texts, such as business documents and reports, ranging from 850 to 1500 words, and then were required to write summaries of the texts.

Before receiving summary instruction, the students were asked to take the reading section of a standardized test, Test of English for International Communication (TOEIC), as a pre-test to identify their English reading proficiency. The maximum score on the TOEIC test for the reading section is 495, and the reliability of the TOEIC test is reported to be 0.90 (Educational Testing Service, 2007). According to the TOEIC pre-test, the 107 students were grouped into an experimental group (N = 54) and a control group (N = 53). The mean score and standard deviation of the 54 students in the experimental group was 287.51 and 33.64 on the pre-test, while the mean score and standard deviation of the 53 students of the control group was 283.39 and 30.16. Significant differences were not determined for the two group’s proficiencies in the pre-test (t(51) = -1.22, p > .01).

System development

The automatic system of keyphrases, three-layer concept map, and scoring, includes the student interface and the teacher interface.

The student interface

Keyphrases are regarded as brief summaries of a text (Erkan & Cicekli, 2007). They may appear in important places (i.e., the title and topic sentences) within a text. To further construct semantic relations between keyphrases to form the three-layer concept maps, Wikipedia Miner served in this study as a referential source containing algorithms, a database package, and an open-source toolkit. The Wikipedia Miner includes algorithms for generating semantic relatedness measures, which quantify the extent to which different words or concepts relate to each other (Milne & Witten, 2013).

For example, as shown in Figure 2, when users entered words, such as soldier, Nationalist army, war, family and junior high school, into Wikipedia Miner to conduct a trial for the calculation of semantic relatedness, the results indicated that the word soldier is 81% related to Nationalist army, 72% related to war, 63% related to family, and 60% related to junior high school. The five keyphrases were correlated to each other in a given text based on the

![Figure 2. Keyphrase extraction](image-url)
semantic relationships of synonyms, hypernyms, and hyponyms. In this study, we developed our algorithms and used Taiwan Panorama’s database to extract and compare word semantic relatedness and help students construct their three-layer concept maps for writing summaries as they read Taiwan Panorama’s articles. With the scaffold of keyphrase extraction, students were able to understand which keyphrases were important to describe the same key concepts in a paragraph. The reliability of the keyphrase extraction in this study is 0.82 (Tsai, Hsu, Yang & Yeh, 2012).

The new three-layer concept map in this study contains three hierarchical and definite layers, with the keyphrases grouped into these three layers. According to keyphrase extraction, this study developed a fill-in-the-blank concept map for students to visualize their knowledge structures (Novak, 2002). The first layer in the center of the concept map is the central idea of the text. The title may become a central idea if it contains high-frequency and related content words and phrases. The second layer of the concept map is the main idea in each paragraph. The numbers of nodes in the second layer of the concept map depends on the numbers of paragraphs in a text. The calculation of the frequency of each verb or noun phrase appearing in the topic sentence as well as in the content of each paragraph is taken into account. The outer third layer is the collection of supporting ideas in each paragraph. As shown in Figure 3, the new concept map is automatically provided by the system after students partially fill in their keyphrases for each paragraph.

Ruiz-Primo and Shavelson (1996) have proposed that automatic scorings of the concept maps can be used by comparing student and teacher maps. The criteria could examine students’ concept structure, semantic content, and their comprehension (Liu & Lee, 2013). After reading an English text, the students’ three-layer concept maps are automatically scored by the system (see Figure 4). There are two purposes for automatic scoring in this study. First, it provides teachers with the opportunity to examine whether their students truly understand the texts or not. Second, it provides students with the opportunity to reflect on what they have understood in the texts and a chance to revise their first draft summaries into final ones.

The first sentence of the summary is calculated by the keyphrases of the first layer (central idea), and the following sentences are calculated by the second (main idea) and third (supporting idea) layers. Students are required to follow the sequence of nodes and links from each paragraph to write their summaries. For example, in Figure 5, the student received immediate feedback about their scoring (80 score points) and guidance after submitting his summary. The scoring of the summary was calculated by 3 keyphrases in central ideas such as “Wang Zhongtian,” “70s” and “calligrapher” *15 score points (45/3=15), and 14 keyphrases in the main idea of the paragraph such as “soldier,”
“Nationalist army” and “fitness officer”*2.5 score points (45/18 = 2.5), which amounted to a total score of 80 points. The boldface words of the summary are keyphrases calculated by the system, and the scoring of keyphrases are based on the automatic map.

The teacher interface

Finding His Calling in His 70s: Calligrapher Wang Zhongtian
Like other old soldiers of his generation, Wang, born in 1926 in Shandong’s Zibo, tells a story that starts with being uprooted amid the chaos of war. From a poor family, he was forced to be a soldier after graduating from junior high school at 18. The following year, he left his family and went alone with the Nationalist army to Taiwan. The silver lining in the cloud was that he was selected by his commanding officer to become a lieutenant.

Student’s map  Automatic map

Summary

Figure 4. Scoring based on the student’s map and automatic map

Finding His Calling in His 70s: Calligrapher Wang Zhongtian
Like other old soldiers of his generation, Wang, born in 1926 in Shandong’s Zibo, tells a story that starts with being uprooted amid the chaos of war. From a poor family, he was forced to be a soldier after graduating from junior high school at 18. The following year, he left his family and went alone with the Nationalist army to Taiwan. The silver lining in the cloud was that he was selected by his commanding officer to become a lieutenant.

Student’s map  Automatic map

Summary

Wang Zhongtian in his 70s became a calligrapher. He was forced to become a soldier in the Nationalist army due to the war. He had been taken on as a fitness officer and finally served as a lieutenant in the military. As he retired, he received a monthly pension and opened a noodle shop to support his family. Later on, he closed his shop and was employed as an orderly by following his wife. After retirement, he joined a calligraphy class, and then changed his life from a retired soldier to a master teacher in calligraphy.

Figure 5. Automatic scoring of the summary
In the teacher interface, teachers can monitor their students’ progress in writing summaries by observing the log files. The log files record students’ actions in the system such as grasping central and main ideas, requesting automatic concept maps, receiving scores to evaluate their summaries, and facilitates teachers to provide individual scaffolding to solve their students’ reading and writing difficulties.

Procedure of data collection

This study incorporated onsite instruction with the automatic system for a period of 18 weeks, including two different versions of TOEIC tests serving as pre- and post-tests. As shown in Figure 6, when the instruction began, the TOEIC pre-test was provided to identify the students’ level of English language proficiency. In the first eight weeks (4 academic texts), the students were guided in how to obtain central ideas from the topic and grasp main and supporting ideas from each paragraph. They were required to delete irrelevant information, generalize each group of main and supporting ideas, and integrate each group of main ideas to write summaries of first drafts with onsite instruction.

After eight weeks of onsite instruction (2 hours per week and 4 academic texts), automatic scaffolding and measurement of three-layer concept maps were introduced to the students of the experimental group for another eight weeks (another 4 academic texts), while the students in the control group continued their onsite instruction, drew paper-based concept maps and received the teacher’s feedback. The students in the experimental group first read academic texts online, received automatic keyphrases and three-layer concept maps, and measured their summaries by automatic scoring. Afterwards, they had to revise the same first drafts into their final drafts. As the students of the experimental group had requested and observed the keyphrases in three-layer concept maps, they could diagnose their summaries to consider whether they had recognized the central and main ideas or not. The students then received automatic scoring to measure their summaries and examine their improvement from the first to final drafts. In contrast, the students of the control group continued their onsite instruction, reading 4 academic texts and learning to draw paper-based concept maps with three layers. The teacher may have provided students with keyphrases when they encountered reading difficulties. The students then revised their first drafts into final ones after receiving teacher feedback for summary revisions and scoring.

At the end of the instruction, the university students were required to take a TOEIC post-test to examine reading improvement between the experimental and control groups. The comparison between first and final drafts examined the students’ progress in writing summaries with a group using automatic scaffolding and measurement of three-layer concept maps and a group using paper-based concept maps. The students’ action details on the request for automatic three-layer concept maps for writing summaries recorded in the system were also analyzed. Finally, the students’ perceptions of using automatic three-layer and paper-based concept maps were revealed by two open-ended questionnaires.

Figure 6. Procedure of experimental design
Procedure of data analysis

Data were analyzed in terms of the students’ pre- and post-test scores, the process data of requesting automatic keyphrases, three-layer concept maps and scoring, the comparison between students’ first and final drafts, and the open-ended questionnaires. First, the pre- and post-tests and the first and final drafts results of the experimental and control groups were compared by means of two paired-sample *t*-tests to examine the students’ progress in reading comprehension. Second, the process data of requesting automatic keyphrases, three-layer concept maps and scoring for improving summaries was explicitly represented to show how the students of the experimental group improved their reading comprehension and summarizing skills based on the automatic scaffolding and measurement. Improvement in reading and writing from the students of the control group was examined by paper-based concept maps combined with teacher feedback. Finally, content analysis was administered to investigate the students’ perceptions of a group using automatic three-layer and a group using paper-based concept maps for writing summaries.

Four stages of content analysis were proposed in this study (Patton, 2002), including coding, categorization, description and interpretation. First, the researcher and the research assistant highlighted meaningful statements in the students’ actions in the log files of the system and two open-ended questionnaires. Afterwards, the meaningful statements were grouped into categories with the students’ perceptions on using the automatic three-layer concept maps and paper-based concept maps for learning to write summaries. Next, the researcher described the statements and summarized the main points. Finally, the researcher interpreted the main points by offering explanations, drawing conclusions and making inferences. The inter-rater reliability (the researcher and the research assistant) was 0.85. The disagreement between these two raters was resolved by discussion. Data interpretation driven by these methods is further explained in the following sections.

Results

To investigate the effects of automatic scaffolding and measurement of three-layer concept maps to improve students’ reading comprehension and summaries, four categories were addressed in this section: (1) the comparison of the reading progress between the experimental and the control groups is provided by the TOEIC pre- and post-tests and the first and final drafts of summaries, (2) the influence of keyphrases and three-layer concept maps on summary writing and reading comprehension (i.e., process data of the students’ three-layer concept maps and first and final drafts), and (3) the students’ perceptions of summary improvement with a group using automatic scaffolding and measurement of three-layer concept maps and a group using paper-based concept maps.

Students’ progress in reading comprehension and summary writing

Two paired-sample *t*-tests compare the mean scores of the pre- and post-tests in reading proficiency for both the experimental and control groups. The results show that the students of the experimental group made greater improvement in their reading comprehension, as their mean scores increased from 283.39 in the pre-test to 328.38 in the post-test. There is a significant difference between the pre-and post-tests in reading (*t*(27) = -4.47, *p* < .00). In contrast, the mean scores of the control group in the post-test (310.59) show that little improvement in reading occurred when compared with the mean score from the pre-test (287.51). A slight difference between the scores of the pre-and post-tests is noted (*t*(23) = -2.83, *p* < .03). Comparing the results of the students’ first and final drafts of the two groups, as shown in Table 1, the findings indicate a significant principal effect for the students of the experimental group in their final drafts (*t*(27) = -8.54, *p* = .00) after receiving automatic scaffolding and measurement of three-layer concept maps. In contrast, the students of the control group made little improvement in their final drafts (*t*(23) = -5.61, *p* = .05) compared with their first drafts after receiving only teacher feedback. The results of reading progress by comparing the two groups’ post-tests and summary improvement indicate that the students of experimental group using automatic scaffolds and measurement of three-layer concept maps to improve their summary writing and reading comprehension were superior to those of the students in the control group of using paper-based concept maps.
Table 1. Comparison between first and final drafts in the two groups (N = 107)

<table>
<thead>
<tr>
<th>Summary</th>
<th>Experimental group</th>
<th></th>
<th>Control group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>First draft</td>
<td>72.05</td>
<td>11.14</td>
<td>-8.54</td>
<td>.00**&lt;br&gt;76.68</td>
</tr>
<tr>
<td>Final draft</td>
<td>84.53</td>
<td>5.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. **p < 0.01. *p < 0.05.

The process of requesting automatic three-layer concept maps for improving summaries and reading comprehension

The onsite instruction was conducted to help the students accumulate the fundamental knowledge necessary for drawing their concept maps and writing summaries. After identifying the main ideas in each paragraph, the students were encouraged to practice filling in keyphrases on a paper-based three-layer concept map as a “planning” stage before writing their summaries. However, the students had difficulties improving their summaries as they neither grasped the main ideas in each paragraph nor received immediate feedback of scoring. In addition, without the log files of process data, the students’ summarizing processes of identifying the main idea were not documented. As a result, the teacher could not monitor the students’ reading and summarizing difficulties.

In online instruction, the automatic scaffolding and measurement of three-layer concept maps to improve summaries and comprehension are described. The online system tracked and recorded students’ concept-constructing processes. In this study, automatic keyphrases and three-layer concept maps demonstrate students how to construct their own maps and help them to clarify the relations in each group of main ideas and make a logical and well-organized summary. For online instruction, student A was one of the sample cases from the experimental group used to examine his summary improvement from first to final drafts after using automatic scaffolding and measurement.

Initially, student A was asked to post his first draft after the onsite instruction, “A question of numbers” in the system. A score was automatically generated and shown on the screen to inform him how good or poor his first summary draft was. He received a score of 68 points out of a full scale of 100 from the automatic system (see Table 2). After student A requested the automatic scaffolding of keyphrases and three-layer concept maps (see Figure 7), student A received a score of 87 points for his three-layer concept maps. The automatic map (see Figure 8) was also shown to clarify the connection between each group of main ideas and supporting details in the text. Based on his own concept maps and the automatic map, he then revised his first draft into a final by extracting central and main ideas again and obtained a score of 91 points in his final draft. That is, the automatic scaffolding of the keyphrases and three-layer concept maps facilitated student A to grasp more important main ideas and necessary details to write the summary. For example, “numbers” was a keyword in this article, but student A did not pay attention to the keyword “numbers” in his first draft. After receiving the automatic map, he discovered that he had to mention “38 neighbors” and “38 witnesses” while revising summaries from the first to final draft.

Table 2. Comparison between student A’s first and final drafts in the system

<table>
<thead>
<tr>
<th>Summary</th>
<th>First draft</th>
<th>Final draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>The key of <strong>bystander effect</strong> is question numbers. It is a complicated human reaction. For example, Genovese was <strong>murdered</strong> but his <strong>neighbors</strong> did not call the police. Psychologists set up an experiment with three groups of students to prove this theory. It's just human nature. The results supported Darley &amp; Latane's theory, and the <strong>bystander effect</strong> can apply to daily situations. However, <strong>bystander effect</strong> happens because of the present of other people.</td>
<td>Summary = 68</td>
<td>Summary = 91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The key of <strong>bystander effect</strong> is question numbers. Proposed by social psychologists, <strong>bystander effect</strong> refers to human complex reaction. For example, Genovese was <strong>murdered</strong> but his <strong>neighbors</strong> did not do anything. Though there were <strong>38 witnesses</strong>, people decreased the chances and <strong>responsibility</strong> to help others. Darley and Latane also set up an experiment to group students into three groups to prove this theory. The results presented the <strong>bystander effect</strong> can be related to large numbers and happen in daily situations. However, <strong>bystander effect</strong> is one of the factors to affect people’s decision.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The students’ perceptions of using automatic three-layer and paper-based concept maps in reading comprehension and summary writing

According to the content analysis of two open-ended questionnaires between the experimental and control groups, all participants in the experimental group stated that the keyphrases and three-layer concept maps assisted them with identifying the central and main ideas for improving their reading comprehension and summaries. All of the participants from the experimental group indicated that the automatic keyphrases really helped them grasp the main
ideas in each paragraph and construct the three-layer structure of concept maps to clarify the logical relations across paragraphs. The automatic map and scoring also helped them understand the quality of their summaries (see Table 3). Most of the participants (93%) confirmed that the scaffold of the automatic map helped them to clarify what important keyphrases were missing and then improve their summaries. However, there were 15 participants (28%) who encountered difficulties grasping the main ideas while reading the passage. They usually reread the text and figured out what main ideas should be included while completing the fill-in-the-blank concept maps. The participants of the control group also agreed that using paper-based concept maps helped them to clarify text misunderstanding and improve reading comprehension. However, 27 participants (51%) claimed that they experienced problems with drawing the paper-based concept maps since they had trouble grasping the keyphrases and clarifying important information and irrelevant details for constructing the maps. Twenty-three participants (43%) proposed that they would like to read the automatic map or more-proficient peers’ maps and compare those with their own for improving reading comprehension and summaries.

Table 3. Students’ perceptions toward online and paper-based concept maps between experimental (N = 54) and control groups (N = 53)

<table>
<thead>
<tr>
<th>Statements</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental group</strong></td>
<td></td>
</tr>
<tr>
<td>1. I can have a basic understanding of and identify the main ideas in each paragraph through requesting automatic three-layer concept maps, and then receive immediate feedback through scoring to evaluate my summaries. I can examine how good my summary is.</td>
<td>54</td>
</tr>
<tr>
<td>2. I like to read the automatic map and more-proficient peers’ map to clarify what important keyphrases I missed.</td>
<td>50</td>
</tr>
<tr>
<td>3. Each layer of the concept map helps me clarify the relationships between the main ideas in each paragraph for writing a good summary.</td>
<td>42</td>
</tr>
<tr>
<td>4. Sometimes I have difficulties grasping main ideas while reading the text. I usually reread the text and figure out what the main points are in a paragraph.</td>
<td>15</td>
</tr>
<tr>
<td>5. With the help of automatic three-layer concept maps, I can select relevant information and delete irrelevant details when writing a summary.</td>
<td>8</td>
</tr>
<tr>
<td><strong>Control group</strong></td>
<td></td>
</tr>
<tr>
<td>1. By drawing paper-based concept maps, I can distinguish the relationships across paragraphs while reading the passage.</td>
<td>53</td>
</tr>
<tr>
<td>2. I think that drawing paper-based concept maps helps me write my summaries.</td>
<td>35</td>
</tr>
<tr>
<td>3. I have difficulties in drawing concept maps because I was unable to grasp the key ideas and was even unable to clarify what important information and irrelevant details are.</td>
<td>27</td>
</tr>
<tr>
<td>4. Though I received teacher feedback, I still wanted to read more-proficient students’ maps to improve mine.</td>
<td>23</td>
</tr>
<tr>
<td>5. I had to wait a long time to receive my teacher’s feedback.</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4. Student perceptions toward writing summaries (N = 107)

<table>
<thead>
<tr>
<th>Statements</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think that the summarizing skill is important as I need to integrate each group of main ideas and represent my understanding of the text.</td>
<td>107</td>
</tr>
<tr>
<td>2. Based on the concept maps, I can construct my summaries without rereading the texts too many times.</td>
<td>93</td>
</tr>
<tr>
<td>3. I think I was unable to use accurate words or sentences to paraphrase the text in writing summaries, but now I can do with the support of the concept maps.</td>
<td>72</td>
</tr>
<tr>
<td>4. I may involve too much detailed information and lack transition words to connect my main points.</td>
<td>24</td>
</tr>
<tr>
<td>5. I have difficulties with how to use different sentence structures and key words to write a good summary.</td>
<td>15</td>
</tr>
</tbody>
</table>
As shown in Table 4, all of the participants (100%) in the experimental and control groups agreed that the summarizing skill was helpful for them to improve their reading and writing skills. Most of the participants (87%) indicated that concept maps helped them identify important keyphrases and examine what ideas were missing or irrelevant before writing their own summaries. However, 72 participants (67%) indicated that they were unable to paraphrase the texts with accurate English words or sentences when writing summaries. They could not completely use their own words to write summaries.

Discussion and conclusion

From the results of this study, it was found that the automatic scaffolding and measurement of three-layer concept maps had a distinct influence on the students’ reading comprehension and summary writing skills (e.g., Sturm & Rankin-Erickson, 2002; Liu, Chen & Chang, 2010). This result is in line with previous studies that concept mapping is able to monitor students’ comprehension and improve their reading accuracy (Redford, Thiede, Wiley & Griffin, 2012) and to externalize their internal language knowledge in self-regulated learning environment (Wang, Peng, Cheng, Zhou, & Liu, 2011). According to keyphrases which were extracted from Taiwan Panorama’s database, this study developed three hierarchical and definite layers to indicate related keyphrases for grasping main ideas in each paragraph to construct a map for writing summaries.

First, compared with the students’ pre-and post-test scores as well as first and final drafts between the experimental and control groups, the students of the experimental group made greater improvement in reading comprehension and summarizing skills after receiving automatic scaffold and measurement of three-layer concept maps. The automatic map was given to the students to aid them in clarifying what important words were missing and unnecessary details were included for improving concept map construction to write summaries. In contrast, the students of the control group made only little progress in reading comprehension and writing summaries since they increased cognitive loads to identify keyphrases and drew paper-based concept maps to write summaries without the automatic feedback from the system.

Second, in the traditional classroom setting, there are no explicit guidance and monitoring tools to help students how to learn writing summaries and record their summarizing processes. To visualize and externalize the students’ mental thoughts while learning to write summaries, implemented by the automatic system of keyphrases, three-layer concept maps, and scoring, teachers were able to examine whether their students truly understood how to construct a map or not and students could reflect on what they had understood of the texts and revise their first drafts of summaries into final ones. In addition, the online learning environment also allowed the students to read more-proficient peers’ concept maps and summaries, which can help them solve their own reading and writing problems. They further learned how more-proficient peers constructed their concept maps and how they write a well-organized summary.

Third, the investigation of the students’ perceptions with a group using automatic scaffolding and measurement of three-layer concept maps and a group using paper-based concept maps to improve summaries and reading comprehension were illustrated. The majority of students in the experimental group agreed that the automatic keyphrases and three-layer concept maps facilitated them to visualize and organize the flow of ideas to structure key concepts as well as improve their reading comprehension. The automatic scoring and feedback provided by the system also helped them understand how good their summaries really were. In contrast, though the students of the control group drew paper-based concept maps and received the teacher’s general feedback to improve the summaries, they still had difficulties in identifying each key concept in the paragraphs.

Some limitations were also found in this study. First, though 107 university students enrolled in this study to investigate automatic keyphrases and three-layer concept maps on EFL university students learning how to write summaries, the results of this study might not be representative of all the problems that EFL university students encounter and all of the solutions which can be solved by using automatic keyphrases, three-layer concept maps and scoring. Second, samples of the summary should be automatically provided by the system in future so that novice students can view good examples when learning to write summaries. Finally, the teacher’s perspectives toward teaching summaries by the automatic three-layer concept maps could be further discussed by interviews. The teachers may play different roles in teaching as they act as a dominator in onsite instruction and a facilitator in online instruction.
Acknowledgements

This article is partially supported by National Science Council in the Republic of China, Taiwan (NSC 101-2410-H-224-029-MY2 and MOST 103-2410-H-224-002).

References


Comparison of Collaboration and Performance in Groups of Learners Assembled Randomly or Based on Learners’ Topic Preferences

Karina L. Cela1*, Miguel Ángel Sicilia2 and Salvador Sánchez2

1Computer Science Department, Universidad de las Fuerzas Armadas – ESPE, Ecuador // 2Computer Science Department, University of Alcalá, Alcalá de Henares (Madrid), Spain // klcela@espe.edu.ec // msicilia@uah.es // salvador.sanchez@uah.es

*Corresponding author

(Submitted July 15, 2014; Revised February 5, 2015; Accepted March 30, 2015)

ABSTRACT

Teachers and instructional designers frequently incorporate collaborative learning approaches into their e-learning environments. A key factor of collaborative learning that may affect learner outcomes is whether the collaborative groups are assigned project topics randomly or based on a shared interest in the topic. This is a particularly important question for adults, whose performance can depend strongly on how closely the project topic relates to their professional goals. In this study involving an on-line course for 103 professionals we divided the learners into two parallel sections that differed in how they would be assigned to groups to perform collaborative tasks on different topics. In one section, learners were assigned randomly into groups and were given task topics randomly as well. In the other section, they were assigned to a group based on whether they shared a common interest in the topic given to that group. We used Social Network Analysis and Content Analysis to assess the level of collaboration in quantitative and qualitative terms (knowledge construction). Both groups showed a low level of knowledge construction and similar levels of centrality degree and learning performance. However, more learners participated in the collaborative tasks if groups had been assigned based on learners’ topic preferences. Our findings suggest that forming groups of adult learners based on topic preferences in on-line environments can increase the number of learners that collaborate, but it does not necessarily improve learner performance.

Keywords

Collaborative learning, Learner preference, SNA, Content analysis, eLearning

Introduction

Designed to mimic the fact that collaboration is an everyday activity, the collaboration approach has been widely used since the 1970s (Strijbos, Martens, & Jochems, 2004), and it has taken on particular prominence in education within the last decade (Laal & Ghodsi, 2012). eLearning environments typically incorporate several tools and communication channels to support collaboration among learners (Koschmann, Kelson, Felteovich, & Barrows, 1996).

The term “collaboration” has been widely applied and studied in many fields with the aim of optimising and exploiting the synergies made possible by collective intelligence (Levy, 2007). Collaboration supports corporate success by fostering the transfer of knowledge within multidisciplinary teams (Skyrme, 2013). Olson, Olson, Carter, and Stororsten (1992) analysed how different members of one organisation collaborated during meetings, and they found that a significant amount of time was spent designing the format of the discussion. Robson and Bennett (2000) found that collaborative modes of interaction between small businesses and their suppliers were critical to increasing turnover.

Collaboration in education has been widely studied (Dillenbourg, Baker, Blaye, & O’Malley, 1995). In one seminal study, Dewiyanita, Brand-Gruwel, Jochems, and Broers (2007) found that group satisfaction with collaborative work varied directly with group cohesion, and that most learners enjoyed working on collaborative projects. So and Brush (2008) found that students who participated more in collaborative learning projects were more satisfied with the learning experience than learners who participated less. Educational technology in general and eLearning in particular exploit a range of platforms and communication tools to harness the learning power of collaboration. Evidence suggests that 3D immersive environments, in which learners interact in a virtual learning world, are particularly effective at catalysing collaboration among learners (Lorenzo, Sicilia & Sánchez, 2012).
One of the key questions when designing collaborative tasks is how to assign learners to collaborative groups. This is particularly important when the target audience is adult learners, since they often exhibit strong preferences about what topics they wish to learn and which learning strategies and tactics they wish to use (Knowles, 1970). Thus designing collaborative tasks according to learner’s preferences may provide strategic insights into instructional design for adult learners in on-line courses.

Surprisingly, little is known about how learner’s preferences influence collaborative learning outcomes in eLearning. The present study adopts the relatively new and underexploited approach of combining content analysis (CA) and Social Network Analysis (SNA) to examine the effects of learner preferences on participation in collaborative projects and on learning performance.

Here we compare collaboration in quantitative and qualitative terms and in terms of learning performance between groups of on-line learners organised randomly and groups containing only members interested in the assigned topic. We wanted to know whether learners who work only on their preferred topics collaborate differently than learners who are assigned topics randomly. To gain insights into how learners collaborate under each type of group composition, we took a combined approach of SNA and CA. These complementary techniques provide insights into, respectively, the quantity and quality of interaction (Erlin, Yusof, & Rahman, 2008; Erlin, Yusof, & Rahman, 2009; Rabbany k., Takaffoli, & Zaïane, 2012).

**Theoretical background**

**Collaborative learning**

Collaborative learning is an instructional approach in which students work in groups to solve a problem (Dillenbourg & Schneider, 1995). In learning environments, collaboration can promote critical thinking (Gokhale, 1995), lead to better learning outcomes than traditional methods under certain conditions (Terenzini, Cabrera, Colbeck, Parente, & Bjorklund, 2001), develop social interaction skills and help students develop a sense of responsibility towards one another (Laal & Ghodsi, 2012), and develop social competencies (Bruffee, 1995) cognitive skills (Lê, 2002) and problem-solving abilities (Blaye et al., 1991).

One of the advantages of eLearning is that it can facilitate learning among students living far apart. Indeed, collaborative learning is an integral part of most eLearning environments, and it has become more prominent as a result of technological advances that support communication via chat, email, and forums. At the same time, any collaborative learning strategies adopted in eLearning should be well focused in order to maximise outcomes and avoid learner drop out that plagues many virtual learning initiatives.

Numerous factors should be taken into account when designing collaborative activities. These factors include the nature of the content, technological support available, and competence of course tutors (Dillenbourg, 1999). These factors can strongly influence the probability that learners will isolate themselves or collaborate, complete the course or drop out (Frankola, 2001).

Another key factor potentially important in learner outcomes is how learners are organised into groups to work on collaborative tasks (Kumar, 1996; Dillenbourg, 1999; Chidambaram, 2005). Instructors often ask themselves, should the groups be assigned randomly or according to some criterion, such as learners’ current knowledge of the assigned topic or preference for that topic?

Despite the importance of this issue, relatively few studies have examined it. Webb, Nemer, Chizhik, and Sugrue (1998) found that below-average students working in heterogeneous groups in which they were grouped with above-average students performed better than below-average students working only with other below-average learners. Popov, Biemans, Brinkman, Kuznetsov, and Mulder (2013) found that collaborative groups comprising culturally homogeneous members were more effective at solving assigned tasks than were culturally mixed groups. Alfonseca et al. (2006) studied the effects of grouping learners according to learner style and concluded that some dimensions of learning styles affect the quality of collaborative work. Savicki, Kelley, and Lingenfelter (1996) found that
female-only groups used more words per message and reported greater satisfaction with the collaborative process than did male-only groups or mixed groups.

It is also possible to organise groups based on learner preferences for topics, such that a group receiving one topic comprises only those learners interested in that topic. To date, this type of group composition has not been studied in eLearning environments, despite its potential importance for adult learners, who are generally autonomous learners, capable of deciding what knowledge or skills will be useful for their lives and therefore what they are interested in learning (Knowles, 1970; Aretio, 1988; Wlodkowski, 1986). Therefore we set out to analyse group composition based on the topic preferences of learners in order to see whether it affects how adult learners interact and how they perform in the learning task overall.

**Social Network Analysis**

Social Network Analysis (SNA) aims to study social relationships among actors in a network, where each actor is represented as a node connected to other actors (Wasserman & Faust, 1994). Actors may be persons, organisations, countries, or communities, and they may be connected through social bonds, kinship, economic interests, or personal interests (Haythornthwaite, 1996).

The ability of SNA to handle large data sets involving complex interactions makes it ideal for analysing eLearning systems (Scott & Carrington, 2011; Cela et al., 2014). In a typical eLearning study, data on interactions among learners and between learners and teachers are continuously gathered during the course. SNA identifies patterns in the data and builds a picture of the social structures operating in the eLearning environment (De Laat, 2002). Learning management systems (LMS) widely used today incorporate a range of communication channels, including forums (Bentivoglio, 2009; Erlin et al., 2008; Gottardo & Noronha, 2012), microblogs (Stepanyan, Borau, & Ullrich, 2010) and wikis (Barth, 2010; Kepp & Schorr, 2009). SNA can analyse the extent of knowledge sharing on all of these channels using graphical and mathematical methods.

SNA has proven effective for evaluating social structures in collaborative learning (Nurmela, Lehtinen, & Palonen, 1999). For example the method has provided an overview of how interactions among learners change over time (De Laat, Lally, Lipponen, & Simons, 2007), how students work in a self-taught course (Laghos & Zaphiris, 2006), how learners and teachers interact in on-line courses (Gottardo & Noronha, 2012), and how high- and low-performing students interact (Dawson, 2010).

**Content analysis**

Content Analysis (CA) makes inferences from texts (Berelson, 1952). CA has been widely used to complement SNA to obtain deeper insights into learners’ interactions in on-line environments (De Laat, 2002; Erlin et al., 2008; Erlin et al., 2009; Yang, Yoo, Lin, & Moon, 2010). Researchers have drawn on CA to develop several models to analyse interactions in eLearning courses. For example, Henri (1992) proposed classifying textual features into five categories, two of which capture critical phases of knowledge construction. Newman, Webb, and Cochrane (1995) proposed an analytical model to measure critical thinking. Their model comprised a set of approximately 46 indicators organised into 10 categories.

Gunawardena’s CA-based model focuses on co-construction of knowledge (Gunawardena, Lowe, & Anderson, 1997). This model stipulates five phases of knowledge co-construction; the learner is expected to pass through the phases progressively. The first phase involves sharing and comparing information; the second phase, discovering and exploring dissonance (inconsistency) among ideas, concepts or statements; the third phase, negotiation of meaning and construction of knowledge; the fourth phase, testing and modifying synthesised or co-constructed knowledge; and the last phase, agreeing on newly constructed meaning and applying it (Gunawardena et al., 1997, p. 412). Each phase comprises several subphases involving such actions as corroborating, stating, asking, and clarifying (Table 1).

Gunawardena’s model has been used in several studies assessing the knowledge-building of learners in on-line environments (Aviv, Erlich, Ravid, & Geva, 2003; De Laat, 2002). Gunawardena et al. (1997) originally developed and validated the model using data on the interactions among 554 undergraduates in a global forum. De Laat (2002)
later used the same model to analyse the online discourse of 46 participants in an online community, while Sing and Khine (2006) analysed knowledge-building among 11 teachers and 1 tutor in an online forum. Zhang and Zhang (2010) studied the discourse of 120 participants and one teacher in a blended learning course. Therefore we chose Gunawardena’s model to analyse collaboration quality in terms of knowledge-building. We combined the conceptual depth of the qualitative Gunawardena model with the quantitative measures of SNA in order to gain the most complete possible picture of network interactions.

Table 1. Phases and subphases of knowledge co-construction in Gunawardena’s model and the corresponding actions or textual indicators*

<table>
<thead>
<tr>
<th>Phase I Sharing/comparing of information</th>
<th>Subphase</th>
<th>Action or textual indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA</td>
<td>Statement of observation or opinion</td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>Announcement of understanding</td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>Corroborating examples</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Making inquiries to clear up subtleties of explanations</td>
<td></td>
</tr>
<tr>
<td>IE</td>
<td>Definition, illustration, or fleshing-out of a problem</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase II Discovery and exploration of dissonance or inconsistency</th>
<th>Subphase</th>
<th>Action or textual indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>II A</td>
<td>Recognising differences</td>
<td></td>
</tr>
<tr>
<td>II B</td>
<td>Asking and answering questions to identify the source and degree of difference</td>
<td></td>
</tr>
<tr>
<td>II C</td>
<td>Adopting the position of the participants and supporting arguments or opinions on the basis of data collected, documentary references, etc.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase III Negotiation of meaning/Co-construction of knowledge</th>
<th>Subphase</th>
<th>Action or textual indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>III A</td>
<td>Negotiating or clarifying the meaning of terms</td>
<td></td>
</tr>
<tr>
<td>III B</td>
<td>Negotiating how arguments should be weighted</td>
<td></td>
</tr>
<tr>
<td>III C</td>
<td>Identifying overlap among conflicting opinions</td>
<td></td>
</tr>
<tr>
<td>III D</td>
<td>Statements embodying compromise or co-construction</td>
<td></td>
</tr>
<tr>
<td>III E</td>
<td>Statements that integrate or accommodate metaphors or analogies</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase IV Testing and modifying proposed knowledge synthesis or co-construction</th>
<th>Subphase</th>
<th>Action or textual indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV A</td>
<td>Testing and verifying proposals</td>
<td></td>
</tr>
<tr>
<td>IV B</td>
<td>Testing against existing cognitive schemata</td>
<td></td>
</tr>
<tr>
<td>IV C</td>
<td>Testing against formal data collected</td>
<td></td>
</tr>
<tr>
<td>IV D</td>
<td>Testing against contradictory evidence in the literature</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase V Agreement about, and application of, newly constructed meaning</th>
<th>Subphase</th>
<th>Action or textual indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA</td>
<td>Summarising agreements</td>
<td></td>
</tr>
<tr>
<td>VB</td>
<td>Applying new knowledge</td>
<td></td>
</tr>
<tr>
<td>VC</td>
<td>Statements that demonstrate participants’ awareness of their own learning and understanding of knowledge</td>
<td></td>
</tr>
</tbody>
</table>


Methodology

Research questions

The goal of this study was to assess the collaboration and performance of learner groups engaged in a collaborative task as part of an eLearning course, and to examine whether these factors depended on whether learners were placed in groups randomly or according to whether they had a preference for the task assigned to the group:

- **Random** condition (control group): Learners were assigned randomly to groups, and groups were randomly assigned topics.
- **Preference-based** condition (intervention group): Learners sharing a preference for a given topic were assigned to the same group, and the group was asked to work on that topic.

These two groups were analysed separately and compared in order to address the following research questions:

- **RQ1**: Does grouping learners by topic preference instead of randomly affect their performance?
• RQ2: Does grouping learners by topic preference instead of randomly affect their collaboration in terms of quantity and/or quality?

Context and participants

This study was conducted in two sections of the same 5-week course entitled “Creation of learning objects using eXeLearning” during the fall semester 2012. This course is one of many optional, free courses offered to professionals and other interested adults by the University of Armed Forces (ESPE) in Sangolquí, Ecuador. The course was offered via the Moodle LMS.

At the start of the course, the 103 participants, all graduate-level professionals from Sangolquí, were randomly assigned into two sections. Both sections were given the same content and activities by the same instructor. Course content was organised into four modules, each of which lasted 1 week, except the third module, which lasted 2 weeks. Module 1 was an introduction intended to ensure that all students shared the same minimal level of computer skills going into Module 2. The first module introduced students to tools of communication and collaborative working, such as internal messaging, forums and wikis. Module 2 presented background on learning objects, while Module 3 allowed participants to practice creating learning objects using eXelearning software. Finally, Module 4 was a practical section in which students worked in groups to configure and publish learning objects in the Moodle LMS. For this final task, groups were asked to create a wiki covering a topic that they were assigned. In the random condition, teacher created groups randomly and assigned them topics randomly as well; in the preference-based condition, teacher assigned topics to groups based on topic preferences that group members had expressed during previous forum discussions.

Students were provided documents to read to help them complete the Module 4 collaborative assignment, and students could access their group’s forum to discuss and debate ideas. The final wiki was graded based on the quality of the content and how well it fulfilled the teacher’s guidelines and expectations, which were provided to students at the start of the course.

The main characteristics of both sections of participants are presented in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Participant characteristics</th>
<th>Random group (control)</th>
<th>Preference-based group (intervention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men, n</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>Women, n</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>Age range (yr), n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-30</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>31-40</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>41-50</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>51-60</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>&gt;60</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Educational level, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate degree (engineering)</td>
<td>79</td>
<td>78</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>No previous knowledge related to the on-line course material, %</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

Data collection and analysis

Content for analysis was taken from the logs of each group’s forum during the Module 4 collaborative task, where members communicated with one another to perform the collaborative task. We analysed the logs using a combination of SNA to capture quantitative characteristics of learner collaboration, and Gunawardena’s (1997) approach to assess quality of collaboration. All data arose from interactions between learners; no information was collected regarding interactions between learners and instructors.
The random and preference-based conditions were compared using the SNA measures of density and centrality degree, as well as Gunawardena’s phases; comparisons were made using Student’s t-test or the Mann-Whitney U test.

Results and discussion

Learning performance

Learning performance in the on-line course in this study was assessed through individual grades on a final test, as well as a final group grade on the wiki project. This approach is similar to that in several previous studies that have used final grades to assess learning performance (Cho, Gay, Davidson, & Ingraffea, 2007; Loomis, 2000; Parker & Gemino, 2001).

The final grade did not differ significantly between the random condition \( M = 7.94, SD = 1.92, N = 53 \) or the preference-based condition \( M = 7.70, SD = 1.73, N = 50 \).

Social network analysis

Centrality degree and density provide an overview of how nodes in a network interact (Wasserman & Faust, 1994; Scott & Carrington, 2011). In our study, the nodes were the students and the paths between nodes were the forum posts produced during the Module 4 collaborative task.

Centrality

The centrality degree of a node is the number of its adjacent nodes (Freeman, 1979): the higher the degree of centrality, the greater the proportion of nodes adjacent to the node in question, and the more access that node has to information. In this way, the centrality degree of a network (centralisation) assesses how balanced is the participation in a network (Borgatti, Everett, & Johnson, 2013). A high centralisation degree indicates that a few actors control or significantly influence on-line interactions in the network. In the extreme case of a “star network,” for example, centralisation is 100%. In a directed network, two types of centrality degree are often calculated: the “out degree,” which assesses the inequality of participants based on sent messages; and the “in degree,” which assesses the concentration of contributions based on messages received. For the random condition, in degree centralisation was \( M = 17.60 (SD = 9.82) \) and out degree centralisation was \( M = 49.80 (SD = 34.78) \); the corresponding values for the preference-based condition were \( M = 17.83 (SD = 18.68) \) and \( M = 60.77 (SD = 32.69) \). Table 3 shows the in and out degrees of centrality for each learner group under both conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group no.</th>
<th>No. members</th>
<th>In centrality</th>
<th>Out centrality</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>1</td>
<td>6</td>
<td>33.33</td>
<td>33.33</td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5</td>
<td>18.75</td>
<td>18.75</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>6.25</td>
<td>100</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>12.5</td>
<td>75</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>80</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>5</td>
<td>31.25</td>
<td>31.25</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>6</td>
<td>16</td>
<td>16</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>5</td>
<td>18.75</td>
<td>18.75</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>5</td>
<td>6.25</td>
<td>100</td>
<td>0.2</td>
</tr>
<tr>
<td>Preference-based</td>
<td>11</td>
<td>5</td>
<td>62.5</td>
<td>6.25</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>5</td>
<td>6.25</td>
<td>100</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>5</td>
<td>0</td>
<td>31.25</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>6</td>
<td>16</td>
<td>64</td>
<td>0.267</td>
</tr>
</tbody>
</table>
Among the groups under the random condition, group 1 showed the highest in degree of centrality (33.3%). This is substantially lower than the 100% of a star network, suggesting that interactions in all the groups in this condition were quite dispersed among the nodes. The groups with the smallest in degrees were 3 (6.25%) and 10 (6.25%), and these groups also had the highest out degrees (100%). This means that only one learner in the group received posts in response to posts that he or she had sent, akin to a star network. In other words there was no reciprocal dialogue among the members of these groups.

Under the preference-based condition, group 11 had the highest in degree of centrality (62.5%), closer to the star-network maximum of 100% than the corresponding maximum under the random condition. The groups with the smallest in degrees were groups 13 (0%) and 17 (0%). An in degree of 0% means that all learners in the group received a message in the forum; in other words, the network was not like a star. Groups 12 and 20 showed the maximum out degree of 100%, while group 11 showed the smallest out degree of 6.25%.

Density

Density expresses the level of connectivity in a network as a proportion of the maximum possible connectivity (Wasserman & Faust, 1994); it is calculated by dividing the number of observed ties by the maximum possible number of ties (Scott & Carrington, 2011). Thus it varies from 0 to 1, where 1 (100%) refers to a situation in which all nodes are interconnected. The network density was significantly different between the random condition (M = 0.23, SD = 0.17) and the preference-based condition (M = 0.53, SD = 0.289).

Density values for individual groups under both conditions are shown in Table 3. Group 1 in the random condition showed the highest density (0.548), while groups 2 and 9 showed the smallest (0.1). Most groups showed densities much smaller than the highest of 0.548. Under the preference-based condition, in contrast, several groups showed high density: Group 11, 0.95; Group 13, 0.75; Group 15, 0.6; and Groups 18 and 19, 0.8.

Quality of collaboration

In order to assess quality of collaboration, forum logs were codified according to Gunawardena’s model in order to assign learners to the appropriate phase and subphase of knowledge construction. Although the unit of analysis was a single forum post, some posts were assigned to multiple phases because they contained statements consistent with more than one phase. This codification was performed twice in order to detect any inconsistencies in the classification of occurrences. Most posts from groups under the random condition were assigned to phases IA -IE. Most posts from the preference-based groups also belonged to phases IA, IB, and ID, but nearly none to phase IE, which corresponds to the definition, description or identification of a problem; and none to phase IC, in which the learner corroborates examples cited by other learner(s).

Groups in both the random and preference-based conditions made one statement belonging to phase IIA, which corresponds to discovering and exploring dissonance or inconsistency among ideas, concepts or statements. Groups in the random condition made two statements belonging to phase IIB, while those in the preference-based condition made none; this subphase involves asking and answering questions to clarify the source and extent of disagreement. Groups under the random condition made more phase IIA contributions than groups in the preference-based condition; this phase includes negotiating or clarifying the meaning of terms. Conversely, the preference-based groups made more phase IIB contributions, which involves proposing and negotiating new statements embodying compromise and co-construction. These same groups made nearly as many phase IIIE contributions, which involve

| 15 | 5 | 18.75 | 50 | 0.6 |
| 16 | 5 | 31.25 | 93.75 | 0.25 |
| 17 | 5 | 0 | 62.5 | 0.5 |
| 18 | 5 | 12.5 | 75 | 0.8 |
| 19 | 5 | 25 | 25 | 0.8 |
| 20 | 5 | 6.25 | 100 | 0.2 |
proposing or integrating metaphors or analogies. Groups did not make Phase IV contributions under either condition, and only one phase V contribution was recorded (in preference-based groups).

Together, these results show that learners under both conditions showed a relatively low level of knowledge construction: 88% of forum posts were in phase I, which involves comparing and sharing information. These results echo those of other studies with participants in other countries and with different ages from our students. In Gunawardena’s study of messages from 554 list subscribers of a global debate, 93% of posts belonged to the first phase. In De Laat’s (2002) study of 46 professionals in Europe, 72% of messages were assigned to phase I, and Zhang and Zhang’s (2010) study of 120 university students in China found 85% of posts in the first phase.

Most forum posts focused on comparing and sharing information, which suggests that these elementary actions are the priority for learners in order to achieve a collaborative task (Sing & Khine, 2006). This finding also suggests the difficulty of achieving a higher level of knowledge in this eLearning environment.

Statistical results

We performed statistical tests to compare group collaboration under the random and preference-based conditions in terms of the SNA measures of centrality degree and density, as well as to compare knowledge construction under the two conditions according to the relative frequency of forum posts assigned to different phases of Gunawardena’s model based on content analysis. Figure 1 shows these data for each group under the two conditions.

![Figure 1. Assignment of forum posts to different phases in Gunawardena’s model of knowledge construction for random and preference-based learner groups (Phase codes are defined in Table 1) (Image)](image)

The assumption of homogeneous variance was accepted based on Lévene’s test for centrality in degree, centrality out degree, and Gunawardena’s phases I-IV. We used Student’s t-test to compare group means for these variables under the random and preference-based conditions. The results showed no significant difference for any of the variables between the two conditions.

Since the assumption of homogeneous variance was not accepted for the SNA measure density, we compared group results under the two conditions using the Mann-Whitney test. The test indicated that mean density under the preference-based condition was significantly higher ($p = 0.009$).

Together, these results suggest that arranging learners into groups based on their preference for the assigned topic may increase the number of learners who get involved in discussions to complete a collaborative task (the SNA measure of density), but it does not necessarily alter the numbers of messages received by learners (centrality in degree), the number of messages posted (centrality out degree), or the overall level of knowledge construction achieved.

In this way, our study extends the literature on how learner characteristics influence the way that they collaborate. Previous work has already emphasised the importance of gender (Savicki et al., 1996), learning styles (Alfonseca et
al., 2006) and below- and above-average performance (Webb et al., 1998). These findings have so far been obtained in adult learners, who adopt a more goal-oriented focus than younger learners (Wolfgang & Dowling, 1981). It would be interesting to see whether similar results are obtained with other types of eLearning courses in other countries or cultures with learners of different educational backgrounds.

Conclusions

In this study, we assess whether grouping learners by topic preference instead of randomly affects their performance (RQ1) as well as the quantity and quality of collaboration (RQ2).

We combined the quantitative power of SNA to examine centrality degree and density of learner-learner interactions, with the richness of CA to assess the level of knowledge co-construction. Our results show that groups who were assembled and assigned topics randomly showed similar centrality in and out degrees as groups who were assembled and assigned topics based on member preferences. However, more members of preference-based groups participated in forum discussions than did members of randomly assigned groups. Nevertheless, groups under both random and preference-based conditions showed similarly low levels of knowledge co-construction according to Gunawardena’s (1997) model. Most learner contributions were limited to sharing and comparing information through the actions of making remarks, presenting opinions, clarifying information, or citing evidence to corroborate other learners’ examples. Our results suggest that taking into account learner preferences when assigning topics for collaborative projects to adults in an eLearning course may increase the number of learners who participate in discussions, but it may not make any difference to the performance achieved. A major limitation of this study is the small population. These results should be verified and extended in studies conducted in other countries with other types of topics, learners, and collaborative tasks.

Acknowledgments

This work was partially supported by SENESCYT of Ecuador and the MAVSEL project (Mining, data analysis and visualization based on social aspects of e-learning), funded by the Spanish Ministry of Science and Innovation (TIN2010-21715-C02-01).

References


An Integrated Contextual and Web-based Issue Quest Approach to Improving Students’ Learning Achievements, Attitudes and Critical Thinking

Han-Yu Sung1, Gwo-Jen Hwang1* and Hong-Sheng Chang2

1Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan // 2Department of Information and Learning Technology, National University of Tainan, Tainan, Taiwan // hanyu.sung@gmail.com // gjhwang.academic@gmail.com // dopas.chang@msa.hinet.net

*Corresponding author

(Submitted July 19, 2014; Revised February 22, 2015; Accepted February 22, 2015)

ABSTRACT

In the era of computer and communication technologies, fostering students’ high-level thinking abilities has become an important educational objective. Engaging students in web information searching to answer a series of questions related to a target issue has been recognized as a helpful approach for promoting students’ thinking processes. In this study, a contextual learning approach is employed in web information searching activities to improve students’ learning achievement, attitudes and critical thinking ability. The participants were divided into an experimental group, which was guided to use Internet resources to complete problem-based learning tasks with the contextual learning approach, and a control group, which learned with the conventional web information searching approach that situated students in a pure web information searching environment to answer questions for the issue to be investigated. From the experimental results, it was found that the experimental group exhibited significantly better learning attitudes, learning achievement, and better critical thinking than the control group, suggesting the effectiveness of the integrated contextual and web information searching approach in terms of helping students use Internet resources to investigate the target issue in depth.

Keywords

Contextual learning, Web information searching, Learning attitude, Critical thinking

Background and objectives

In this technology-rich era, people need to face new information, knowledge and problems every day. Consequently, they need to have the abilities to organize new and previous knowledge as well as to cope with the problems encountered (Chu, Hwang, Tsai, & Chen, 2009). Several studies have shown that the abilities of “using technologies and information” and “engaging oneself in independent thinking and problem solving” are helpful to students in terms of fostering their critical thinking competence (Johnson, Archibald, & Tenenbaum, 2010; Şendağ & Odabaşı, 2009). Critical thinking is a form of higher order thinking that people use to solve problems encountered in the environments in which they are situated. It consists of three aspects, that is, skills, mental process and procedures. Ennis (1985) indicated that “critical thinking is reasonable reflective thinking focused on deciding what to believe or do.” Educators have indicated that critical thinking is becoming increasingly important owing to the rapid advancement and growth of information and communication technologies (De Leng et al., 2009). Therefore, it is important to develop learning strategies that situate students in an environment of “using technologies and information” and which engage them in independent thinking and problem-based learning tasks (Kuo, Chen, & Hwang, 2014).

Among various instructional strategies, Web-based Issue Quest (WIQ) has been recognized as an effective way of fostering students’ abilities of searching for, selecting, abstracting and summarizing the information on the web to answer a series of questions regarding a particular issue (Harskamp & Suhre, 2007; Hwang, Kuo, Chen, & Ho, 2014; Lo, 2009). It is an instructional method that guides students to learn through a procedure of seeking information on the web to complete problem-based learning tasks (Hwang & Kuo, 2011). It emphasizes the provision of “a series of questions related to a specified issue” during the learning process, such that students are guided to investigate the issue via searching for information and exploring the ways of finding answers to the questions.

However, previous studies have mainly focused on engaging students in searching for information on the web to investigate specified issues, while the real-world contexts related to their daily life experiences have been ignored. Without connecting what they have acquired from the web to what they have experienced in their daily life, students’
learning attitudes and learning performance could be significantly affected (Brown, Collins, & Duguid, 1989; Heath, 1992; Wu, Hwang, & Tsai, 2013). Therefore, it has become an important issue to develop a learning strategy that integrates web-based issue quest activities with practical applications or daily life experiences.

Contextual learning is considered as an instructional method that engages students in a learning process by connecting the learning tasks with real-life situation issues or with contexts that guide them to seek meanings; moreover, it provides students with a context for clarifying how things are correlated or interrelated (Johnson, 2002; Zheng, Perez, Williamson, & Flygare, 2008). During a contextual learning activity, the students are guided to construct meanings based on their own experiences related to the issues proposed by the teachers. Such learning activities may enhance not only the students’ learning outcomes but also their critical thinking.

In this study, an integrated contextual and web-based issue quest approach is proposed to instruct and guide students to investigate the issues raised by teachers and find answers from both the web and real-world environments; moreover, a web-based issue quest environment has been implemented based on the approach. To evaluate the effectiveness of the approach, the following research questions are investigated in this study:

1. Does the contextual learning approach improve the students’ learning achievements in comparison with the conventional approach in the WIQ environment?
2. Does the contextual learning approach improve the students’ critical thinking ability in comparison with the conventional approach in the WIQ environment?
3. Does the contextual learning approach improve the students’ learning attitudes in comparison with the conventional approach in the WIQ environment?

Literature review

Web-based issue quest

Web-based issue quest (WIQ) refers to the learning activities that engage students in utilizing web information in an organized and meaningful manner for investigating or answering a series of questions related to a specified issue (Cheng, Liang, & Tsai, 2013; Mason, Ariasi, & Boldrin, 2011; Tsai & Tsai, 2010), and has been recognized as an essential skill for adapting to the environment (Geçer, 2014). In a WIQ activity, the students are asked to collect data from the Internet using search engines in order to answer questions raised by their teachers, or to state their opinions regarding a particular issue (Polly & Ausband, 2009).

In recent years, researchers have attempted to conduct web-based learning activities to foster students’ higher order thinking ability (Hwang, Tsai, Tsai, & Tseng, 2008; Kim & Hannafin, 2011; Wu, Hwang, Kuo, & Huang, 2013). For example, Kanuka, Rourke and Laflamme (2007) reported several advantages of WIQ activities, including engaging students in structuring the collected information and providing them with the opportunity to identify their roles and responsibilities during the learning process. Kuo, Hwang and Lee (2012) conducted an elementary school science course which further showed that the middle- and low-achievement students who learned with the WIQ approach gained significant benefits in comparison with those who learned by way of traditional instruction.

Moreover, researchers have pointed out the importance of providing learning supports or leading in learning tools to help students collect information on the web for solving problems (Cothey, 2002; Lazonder, Biemans, & Wopereis, 2000). For example, Kuiper, Volman, and Terwel (2005) indicated that students need support when searching for information on the Web. They also pointed out that future studies should focus on the effects of WIQ approaches on the development of students’ deep and meaningful knowledge. Therefore, it has become an interesting and challenging issue to develop and provide learning approaches to engage students in effective WIQ tasks.

Contextual learning

Contextual learning refers to the learning strategies that situate students in the scenarios related to the learning objectives or tasks. It originates from the constructivist theory which emphasizes that effective learning takes place when information is presented in a way that enables students to construct meanings based on their own experiences. Hull (1993, p. 41) indicated the basic assumption of contextual learning and teaching which is that “the mind
naturally seeks meaning in context—that is, in the environment where the person is located—and that it does so through searching for relationships that make sense and appear useful.”

Many educators have pointed out the importance of situating students in authentic contexts and encouraging them to perceive problems, actively collect data, and develop the way to solve problems. Moreover, extending the scope of learning from the classroom to the wider society makes learning more meaningful and helpful to the students (Chu, 2014; Sadler & Zeidler, 2005; Wu & Tsai, 2007). For example, the study of Ogata and Yano (2004) showed that situating students in corresponding real-world contexts could improve their language learning performance; Shih, Tseng, Yang, Lin and Liang (2012) also reported a similar finding based on an experiment of incorporating real-world contexts into a Chinese poetry learning activity. Hwang, Wu, Zhuang and Huang (2013) further reported the benefits of contextual learning for reducing students’ cognitive load and improving their learning achievement in a social studies course. Via bridging the real-world contexts and the learning content, contextual learning activities aim to promote students’ learning interest and motivation, such that they can realize the value of the learning content and apply what they have learned to their daily lives (Hung, Yang, Fang, Hwang, & Chen, 2014).

Lewis and Leach (2006) indicated that students’ ability to engage in reasoned discussions of scientific and technological applications is highly influenced by the key issue, recognizing ability, which requires understandings of the relevant science. Several researchers have pointed out the importance of bringing daily life experience, epistemological and social considerations into instructional settings in schools (Albe, 2008; Kolstø, 2006; Tytler, 2012; Wu & Tsai, 2007). It is expected that students can engage in dealing with the issues by connecting their daily life experiences to what they have learned in school with greater learning interest. Wolfensberger, Piniel, Canella, and Kyburz-Graber (2010) further emphasized that such contextual learning activities not only benefit students, but also help the teachers to make reflections on their instructional design.

However, previous studies mainly used contextual learning as a strategy for traditional instructional or technology-assisted learning settings (Johnson, 2002; Mooij, 2007), while integrating real-world contexts into web information searching activities is rarely found. Consequently, in this study, an integrated contextual and web-based issue quest approach is proposed by including comparing, debating, challenging, and clarifying the findings based on the information found on the Internet, observations of science phenomena, and experiences in the students’ daily lives. Researchers have indicated that such web-assisted learning strategies that promote interactions between peers, in particular, making comparisons and critiques, are able to engage students in critical thinking (Guiller, Durndell, & Ross, 2008; Yang, Newby, & Bill, 2008). With the contextual and web-based issue quest approach, students can share their findings, make critiques of those reported by their peers based on the information collected from the web, and make reflections based on their daily life experiences and the science phenomenon observations. It is expected that, with the help of the proposed approach, the students’ learning achievement, critical thinking ability and learning attitudes can be improved.

**Research design**

The proposed learning approach is shown in Figure 1. It integrates three inseparable elements, that is, the “Problem-based learning mechanism” that guides the students to identify the target issue in depth via conducting a WIQ activity, the “Web-based learning environment” that enables the students to extend their knowledge via searching for information on the web, and the “Contextual Learning Approach” that guides the students to link the target issue to their daily life experiences. A four-step guideline was used to help the students incorporate the contextual learning approach into the WIQ activity:

1. **Web information searching:** Search for information based on the questions of the target issue.
2. **Issue and question identification:** List the information items derived from the web for individual questions.
3. **Contextualization:** For each information item on the list, find an example from daily life experiences by taking photos or collecting data from neighbors, family or peers.
4. **Brainstorming:** Acquire additional relationships among the information items and the daily life examples via discussing with group members online.

For example, for the “electric power and environmental resources” issue in an elementary school social studies course, the teachers might raise several questions:

1. How many nuclear power plants are there in Taiwan? Where are they located?
(2) What is the scientific principle of using nuclear power?
(3) What are the advantages and disadvantages of nuclear power?
(4) Do you agree with developing nuclear power? Why or why not?

In the first step of the WIQ activity, the students are asked to search for information on the web to answer the questions. They need to select proper web pages, abstract the relevant content, summarize the content and express their own opinions based on what they have found on the web. The contextual learning strategies are then used to encourage the students to find daily life examples by taking photos or collecting data from neighbors, family or peers. Those daily life experiences and findings are presented on the web-based learning system to share with peers, who are allowed to comment on the shared content. Following that, a brainstorming activity is conducted in the classroom to encourage the students to share and discuss their findings, daily life experiences and opinions with their group members based on the follow-up issues raised by the teachers (e.g., “What should people learn about electricity resources?” and “What can we do to protect the earth?”). After the discussion, each group of students is asked to summarize their findings and opinions; moreover, they need to present their findings and conclusions to their classmates using a PowerPoint file. The students in other groups are allowed to raise questions and challenge the presentation content after listening to the presentation.

![Web-based Learning Environment](image)

**Figure 1.** Web-based issue quest with contextual learning

**Participants**

The participants were 48 fifth graders from two classes. One class was assigned to be the experimental group \((N = 24)\), and the other was the control group \((N = 24)\). The students in both groups were divided into six learning groups, each of which had four members. The students in the experimental group were instructed and guided to participate in the WIQ activity with the contextual learning, while those in the control group were instructed and guided to participate in the activity with the conventional approach.

**Experimental procedure**

Figure 2 shows the experimental procedure, which consists of three stages, that is, conducting the pre-tests and the pre-questionnaire, introduction to the tools and learning tasks, conducting WIQ activities, and conducting the post-tests and the post-questionnaire.
Fifth graders (N = 48)

Experimental group (N = 24)  Control group (N = 24)

Critical thinking measure, learning attitude questionnaire, the natural science course pre-test  1 week

Introduction to the web-based learning environment  1 week

Contextual web-based issue quest  Conventional web-based issue quest  4 weeks

Critical thinking measure, learning attitude questionnaire, the natural science course post-test  1 week

Figure 2. Experimental procedure

In the first stage, all of the students took the critical thinking test, the learning attitude questionnaire and the pre-test. In the second stage, the students in both groups were instructed with the tools and tasks of the WIQ activity.

Following the instruction, a four-week long learning activity was conducted, as shown in Table 1. Both groups of students were asked to search for information to answer a series of questions related to the issue specified by the teachers. They were also encouraged to discuss with their group members before presenting their findings. Moreover, the students were encouraged to give comments to other groups and make reflections based on the comments from their peers.

Table 1. Illustrative example of a WIQ learning task

<table>
<thead>
<tr>
<th>Issue</th>
<th>Web-based Issue Quest activities</th>
</tr>
</thead>
</table>
| Nuclear power | - Information searching, selecting, abstracting and summarizing  
The teacher asks each group of students to search for information on the web to answer the following questions:  
(1) How many nuclear power plants are there in Taiwan? Where are they located?  
(2) What is the scientific principle of using nuclear power?  
(3) What are the advantages and disadvantages of nuclear power?  
(4) Do you agree to develop nuclear power? Why? |
|       | - In-Group discussion and presentation  
In addition to answering the questions, the students are asked to discuss what they have found. Moreover, they need to summarize their conclusions and make a presentation.  
- Comparing, debating, challenging and clarifying findings  
Students give comments to other groups based on what they have collected. In this stage, students exchange information and make reflections, which helps them to reorganize their knowledge. |

During the learning activity, the students in the experimental group were guided with the contextual learning strategy; that is, in addition to the information from the web, they were guided to understand the issue from different aspects via observing the relevant events or phenomena in their daily life experiences, and connecting the observations to the questions they were dealing with.

On the other hand, the students in the control group were guided with the conventional web-based issue quest approach; that is, they were instructed by the teacher with PowerPoint and videos to realize the theories and
knowledge concerning the target issue, and were asked to answer the same set of questions in the WIQ activity. After collecting data from the Internet and summarizing the findings, each group of students was asked to discuss and present their findings and conclusions.

In the final stage, the students took the critical thinking test, the learning attitude questionnaire and the post-test.

**Measurements**

Both the pre-test and the post-test consisted of 25 multiple-choice questions with a total score of 100. The pre-test was designed to evaluate the students’ basic knowledge concerning the subject unit, such as “Which of the following items does not need electric power? (a) Television (b) Washing machine (c) Air conditioner (d) Radio (e) None of the above” and “What is the most frequent way of producing electric power in Taiwan? (a) Coal power (b) Nuclear power (c) Water power (d) Wind power.” The post-test was a detailed test of the subject unit, such as “In most advanced countries, there is nuclear power. What is the advantage of using nuclear power?” “Why are the people in Taiwan hesitating to build new nuclear power plants?” Both the pre-test and the post-test were developed by consulting two teachers who had taught the course for more than five years.

The learning attitude questionnaire is provided in the appendix. It consisted of 25 items (e.g., “I think taking the course is important” and “I will actively search for more information related to this course”) with a five-point rating scheme. It was modified from the questionnaire developed by Wang, Chu and Hwang (2010) with the help of two experienced teachers to make the items understandable to the students. The Cronbach’s alpha value of the questionnaire is 0.91.

The critical thinking measure was developed by Yeh (2003, 2009) based on the measure proposed by Ennis, Millman and Tomko (1985). The Cronbach’s alpha value of the measure is 0.76. It consists of five dimensions (i.e., recognition of assumptions, inductions, deductions, interpretations, and evaluations of arguments), each of which includes five items, such as “when encountering some problems, I will try to figure out the reason and find a way to cope with the problems,” “When someone else raises a question, I will seriously think about the things related to the question” and “When discussing with peers, I will try my best to understand and listen to those opinions that are opposite to mine.” The full score of the measure is 25. A higher score represents better critical thinking ability.

**Web-based learning environment**

A web-based learning environment, which consists of a “web-based issue quest” system and a “contextual learning and brainstorming” system, is employed to assist the teachers in conducting the WIQ activities by proposing a series of questions regarding a special issue (Hwang, Tsai, Tsai, & Tseng, 2008; Tseng, Hwang, Tsai, & Tsai, 2009). The “web-based issue quest” system is called Meta-Analyzer, and has been recognized by researchers as an efficient tool for conducting WIQ activities (Hwang & Kuo, 2011; Tsai, Tsai, & Hwang, 2011). It consists of a search engine, a question-and-answer interface for students and a learning management interface for teachers. Figure 3 shows the student interface of Meta-Analyzer, which consists of four operation areas: the “question and answer” area is located on the left of the screen, the “information-searching” area is located on the upper right-hand side, and the searched web pages are presented on the lower right-hand side.

Moreover, a “Bookmark management” function is provided, which enables the students to mark the web pages that are highly relevant to the problems they are dealing with. This function not only helps the students to remember some important web pages for revisiting, but also assists the teachers in examining the students’ ability to identify relevant and important web contents. In some web-based learning activities, “marking the most relevant web pages” could be one of the learning tasks asked by the teacher. Via checking the relevance of the web pages marked by the students to the issue to be investigated, the teacher can evaluate if the students’ judgments are correct. Accordingly, teachers can score individual students’ ability by computing the ratio of their “good” and “bad” judgments. The relevance degree of each of their judgments can also be taken into account.

To answer the questions prepared by the teachers, the students can search for information on the web by entering keywords to the information-searching area and browsing the searched web pages that are relevant to the questions.
The search engine will list the links and titles of the searched web pages. The students can click on the links to browse the web pages, copy and paste the relevant contents to the answer area, summarize the collected data, and then submit the answer.

Figure 3. The student interface of Meta-Analyzer

Figure 4. The contextual learning and brainstorming system

The “contextual learning and brainstorming” system was developed using Google Sites, which is a structured collaborative system-development tool offered by Google. It allows students to share daily experiences related to the target issue via presenting photos with group members, and hence engage them in discussions and brainstorming.
Figure 4 shows an illustrative example of group discussions of electric power costs and the possible ways of energy saving and carbon reduction in our daily lives.

Results

Learning achievements

Before participating in this learning activity, the students from both the experimental and the control group took a pre-test to evaluate their basic knowledge. The means and standard deviations of the pre-test scores were 77.04 and 17.66 for the experimental group, and 75.13 and 13.02 for the control group. A t-test performed on the result of the pre-test scores showed no significant difference between the two groups with t = 0.43 (p > .05), implying that the two groups of students had statistically equivalent basic knowledge before participating in this learning activity.

After the learning activity, the two groups of students took the post-test. Table 2 shows the Analysis of Covariance (ANCOVA) result of the post-test scores of the two groups, and the means and standard deviations of the post-test scores, which were 85.58 and 11.87 for the experimental group, and 80.25 and 11.57 for the control group. It was found that the post-test scores of the two groups were significantly different with F = 5.60 (p < .05), implying that the contextual learning strategy was helpful to the students in terms of improving their learning achievement in the social science course; moreover, the adjusted mean of the experimental group’s post-test scores (84.96) is statistically higher than that of the control group (80.88). Consequently, it is concluded that integrating the contextual learning strategy was helpful to the students in terms of improving their learning achievement in the social studies course.

| Table 2. ANCOVA result of learning achievements on the post-test scores of the two groups |
|---------------------------------|----------|-----|----------|-----|-----|
| Group                          | N | Mean | SD | Adjusted mean | F |
| Experimental group             | 24 | 85.58 | 11.87 | 84.96 | 5.60* |
| Control group                  | 24 | 80.25 | 11.57 | 80.88 |     |

Note. *p < .05.

Critical thinking

To investigate the effect of integrating the contextual learning strategy in WIQ activities on students’ critical thinking ability, the critical thinking test scores of the students in the experimental group were compared with those of the control group students based on the critical thinking measure developed by Yeh (2003). The means and standard deviations of the critical thinking pre-test scores were 10.92 and 3.59 for the experimental group, and 10.58 and 2.67 for the control group. The t-test result showed no significant difference on the critical thinking pre-test scores of the two groups with t = 0.37 (p > .05), implying that the two groups of students had equivalent critical thinking ability before participating in the learning activity.

After the learning activity, the two groups of students took a critical thinking post-test. Table 3 shows the ANCOVA result of the post-test scores of the two groups. It was found that the post-test scores of the groups were significantly different with F = 5.06 (p < .05); moreover, the adjusted mean of the experimental group’s post-test scores (13.66) is higher than that of the control group (12.05), showing that the students in the experimental group exhibited significantly better critical thinking ability than those in the control group. Consequently, it is concluded that integrating the contextual learning strategy in WIQ activities is helpful to the students in terms of improving their critical thinking ability.

| Table 3. ANCOVA result of the critical thinking skill post-test scores of the two groups |
|---------------------------------|----------|-----|----------|-----|-----|
| Group                          | N | Mean | SD | Adjusted mean | F |
| Experimental group             | 24 | 13.75 | 3.59 | 13.66 | 5.06* |
| Control group                  | 24 | 11.96 | 2.67 | 12.05 |     |

Note. *p < .05.
Learning attitudes

Before participating in this learning activity, students from both the experimental and the control group completed the “learning attitude scale” to understand whether the two groups of learners had homogeneous attitudes toward taking the course. The means and standard deviations of the pre-questionnaire ratings were 97.42 and 14.01 for the experimental group, and 94.04 and 11.00 for the control group. The t-test result showed no significant difference between the pre-questionnaire ratings of the two groups with \( t = 0.93 \) \( (p > .05) \), showing that the two groups of students had statistically equivalent learning attitudes before participating in the learning activity.

After the learning activity, to understand the differences between the outcomes achieved by applying different teaching methods, the two groups of students completed the learning attitudes post-questionnaire. Table 4 shows the ANCOVA result of the post-questionnaire ratings of the two groups. It was found that the ratings of the two groups were significantly different with \( F = 4.67 \) \( (p < .05) \). The adjusted mean of the experimental group (4.03) was significantly higher than that of the control group (3.74), implying that the contextual learning strategy had a significant impact on improving the students’ learning attitudes toward the social studies course.

| Table 4. ANCOVA result of learning attitudes on the post-test scores of the two groups |
|---------------------------------|--------|--------|-----------------|--------|
| Group                      | N     | Mean   | SD    | Adjusted mean  | \( F \) |
| Experimental group         | 24    | 4.07   | 0.55  | 4.03            | 4.67*  |
| Control group              | 24    | 3.70   | 0.52  | 3.74            |        |

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

Number of web pages browsed, marked and adopted

From the system logs provided by Meta-Analyzer, it was found that the students in the experimental group on average tended to spend more time searching for data from different web pages, while most of the students in the control group answered the questions after finding only one or two relevant web pages, as shown in Figure 5. This finding conforms to their feedback on the learning attitude questionnaire; that is, after participating in the integrated contextual and WIQ activity, the students in the experimental group felt that the course was important, and were more willing to search for more related information.

Discussion and conclusions

This study investigates the effects of integrating the contextual learning strategy in web-based issue quest activities on students’ learning achievement, attitudes and critical thinking ability. A learning activity has been conducted to compare the learning performance of the students who learned with the proposed approach and that of the students who learned with the conventional WIQ approach. The experimental results show that the innovative approach improved not only the critical thinking ability of the students, but also their learning achievements and attitudes. In
addition, the students who learned with the integrated contextual and WIQ approach showed significantly better learning attitudes toward the social studies course than did the control group. The system logs confirmed this finding since the students in the experimental group on average tended to spend more time searching for data from different web pages than the control group students. The finding also conforms to those reported by previous studies that have demonstrated the importance of linking learning to real-world problems or daily life experiences (Chiou, Tseng, Hwang, & Heller, 2010; Hwang, Chu, Lin, & Tsai, 2011).

Several previous studies have indicated that issue-based discussions can promote students’ critical thinking performance since students need to think in depth and consider the issue from different aspects during the learning process (Guiller, Durdell, & Ross, 2008; Şendağ & Odabaşı, 2009; Yang, Newby, & Bill, 2008). However, when conducting issue-based collaborative learning tasks, it is challenging for teachers to engage students in peer discussions and interactions, particularly in most Asian schools where transferring knowledge to students for exams is considered as the main objective of teaching (Lun, Fischer, & Ward, 2010; Stapleton, 2011). Contextual learning has been recognized as a potential strategy to cope with this problem (Connor & Killmer, 2001). Via combining the web-based collaborative issue quest tasks with the students’ daily life experience by asking them to find evidence from daily life contexts, the students’ interest in discussing the issues with their peers was promoted; in the meantime, they were trained to think from different aspects based on the collect evidence to answer the questions raised when discussing the issues. That is, the students were situated in tasks that encouraged them to think deeply and diversely. This could be the reason why the contextual learning approach was able to improve the critical thinking skills of the students in the WIQ activities, and hence the students’ learning achievements were improved as well. Such a finding conforms to those reported by previous studies, namely that engaging students in critical thinking helps them to improve their learning performance (Guiller, Durdell, & Ross, 2008; Yang, Newby, & Bill, 2008).

To the best of our knowledge, the previous relevant studies mainly focused on engaging students in searching for information on the web for investigating the target issue, while students’ daily life contexts were not taken into account. The contributions of this study are to propose an integrated contextual and Web-based issue quest approach and to develop a learning system to support contextual WIQ activities. With the new approach, the investigated issue can be aligned with students’ daily life experiences, thus improving their learning attitude and performance. For those teachers who intend to use the approach in other subjects, the following preparations are needed before engaging students in the contextual WIQ tasks (e.g., web information searching, issue and question identification, contextualization and brainstorming): (1) Select a target issue related to daily life experiences from the course content; (2) Develop a series of questions related to the issue to engage students in searching for information on the web, abstracting web content, summarizing their findings and presenting their opinions based on the findings; and (3) Develop a set of follow-up issues after the students have completed the contextual WIQ tasks.

On the other hand, some limitations to the present study need to be noted. For example, the findings were based on an experiment involving just 48 students; therefore, generalizations about students’ critical thinking ability, learning attitudes and learning achievements could be difficult to make based on such a small sample size. Moreover, the present learning activity was conducted in an elementary school social studies course. To generalize the findings to other courses (e.g., Physics, Chemistry or Geosciences) and samples (e.g., high school or college students), more studies are needed. In the future, we plan to conduct more learning integrated contextual and WIQ activities in different science courses to investigate their effectiveness from different aspects, such as their effects on the learning performance of students with different achievement levels or learning styles.

Acknowledgements

This study is supported in part by the National Science Council, Taiwan, under contract numbers NSC 101-2511-S-011-005-MY3 and NSC 102-2511-S-011-007-MY3.

References


Tsai, P. S., Tsai, C. C., & Hwang, G. J. (2011). The Correlates of Taiwan teachers’ epistemological beliefs concerning Internet environments, online search strategies, and search outcomes. *The Internet and Higher Education, 14*(1), 54-63.


**Appendix**

**Learning attitude questionnaire**

1. I feel that the course content is rich.
2. I think the learning content is very useful to me.
3. I paid full attention to the learning activity in this course.
*4. I often do my own thing in class.*
5. I am interested in this course.
*6. I feel that the course is boring.*
7. I think the difficulty of the learning content meets my knowledge level.
8. I think I have learned a lot in this course.
*9. I don’t like this course.*
*10. The learning content of this course is irrelevant to my future life.*
11. I always actively complete the homework of this course without being reminded.
12. I try to overcome the learning problems encountered in this course.
13. I hope that I can get a good grade on this course.
14. I try to apply what I have learned in this course to my daily life.
*15. I hate to think and infer when doing the homework for this course.*
16. I actively ask the teacher about the learning content that interests me in the course.
17. The course helps me learn to deal with problems in my own way.
18. I prepare for lessons before the class and review what I have learned after the class.
19. Now I pay more attention to observing what happens in my daily life after taking the course.
20. When the teacher is going to start a new unit in this course, I would read the unit content first.
*21. I usually do homework or prepare for the test for this course when the deadline is approaching.*
22. If I do not understand the learning content of this course, I would ask the teacher or discuss it with my peers.
23. I am able to memorize several important terms to remind me of the key points in individual course units.
24. Even if the learning content is difficult to comprehend, I would like to remember it.
25. Sometimes I try to raise and answer questions to evaluate if I fully understand the learning content of this course.

(*reverse question)
Pedagogically-Driven Ontology Network for Conceptualizing the e-Learning Assessment Domain

Lucila Romero1*, Matthew North2, Milagros Gutiérrez3 and Laura Caliusco3

1Faculty of Engineering and Water Sciences, Universidad Nacional del Litoral, Santa Fe, Argentina // 2The College of Idaho, Caldwell ID, USA // 3CIDISI Research Center–UTN–Facultad Regional Santa Fe, Argentina // lucila.rb@gmail.com // mnorth@washjeff.edu // mmgutier@frsf.utn.edu.ar // mcaliusc@frsf.utn.edu.ar

*Corresponding author

(Submitted December 11, 2013; Revised September 22, 2014; Accepted October 24, 2014)

ABSTRACT

The use of ontologies as tools to guide the generation, organization and personalization of e-learning content, including e-assessment, has drawn attention of the researchers because ontologies can represent the knowledge of a given domain and researchers use the ontology to reason about it. Although the use of these semantic technologies tends to enhance technology-based educational processes, the lack of validation to improve the quality of learning in their use makes the educator feel reluctant to use them. This paper presents progress in the development of an ontology network, called AONet, that conceptualizes the e-assessment domain with the aim of supporting the semi-automatic generation of assessment, taking into account not only technical aspects but also pedagogical ones.

Keywords

Ontology, Ontology network, e-Learning, e-Assessment

Introduction

It is well known that e-learning has become a popular and growing business, thus, the number of courses available on Internet is growing rapidly. Among factors affecting e-learning success, proper feedback mechanisms are integral. As Dewey (1963) argues, learning is fundamentally social in nature. Thurmond, Wambach, Connors and Frey (2002) state that diversity in assessment and perception of interaction with others influences e-learning satisfaction considerably. The use of different evaluation methods in an e-learning system causes learners to establish a connection between themselves and their instructors, which in turn helps to ensure that their learning efforts are properly assessed (Sun Tsai, Finger, Chen, & Yeh, 2008). There are a variety of instruments for assessment; however, in an e-learning system, only a few of them are generally used, with the most popular being objective instruments such as multiple and simple choice.

In the area of e-assessment, there are two problems to overcome. First, e-assessment must be accepted by educators. A tool to support devising and creating valid and reliable assessments, from a pedagogical perspective, is needed. This means that we must define a mechanism to validate whether the assessment covers all of the learning objectives of a course and satisfies certain pedagogical principles such as those proposed by Bolivar (2011). E-learning systems must allow educators create diagnostic, summative, and formative assessment, along with self-assessment, co-assessment, and hetero-assessment.

In the e-learning context, ontologies, and web semantic technologies are widely used with different purposes. Chang and Chen (2009) propose a tool for peer assessment to satisfy the requirements of cooperative learning in an e-learning environment. Gladun, Rogushina, García-Sanchez, Martínez-Béjar, and Fernández-Breis (2009) propose a domain ontology to assess learners’ skills. In this case, the domain ontology is not only the learning instrument but also a means for testing and teaching. Also, in research literature, different approaches that define ontology as a structure to guide the automated design of assessment can be found in (Castellanos-Nieves, Fernández-Breis, Valencia-García, Martínez-Béjar, & Iniesta-Moreno, 2011; Simperl, Mochol, Burger & Popov, 2010). However, most of these approaches are based on individual ontologies that only model a part of the assessment domain, considering only the structural aspects of the assessments. For example, Castellanos-Nieves et al. (2011) defined an ontology for supporting open-question generation, whereas Simperl, et al. (2010) model only true-or-false-type questions. Related to assessment in an e-learning environment, Frutos-Morales, Sanchez-Vera, Castellanos-Nieves, Esteban-Gil, Cruz-Corona, Prendes- Espinosa, and Fernandez-Breis (2010) propose the use of semantic web
technologies to generate semantic feedback in the TL process. Despite the efforts made in this sense, much work remains, especially when considering the pedagogical aspect of the e-learning environment.

Farrús and Costa-Jussà (2013) propose the generation of assessments on demand, where assessment is composed only of open questions, the authors do not specify the ontologies used in that process. In Rupere, Chaka, Zanamwe and Mavhumwa (2013), authors propose the use of language ontology in web system, which lets assess students knowledge through short open questions. In Elsayed, Eldahshan, and Tawfeek (2013), authors propose to assess learners with multiple-choice questions and open questions related to the area of mathematics that require a variable number of steps to reach the solution using semantic systems. Fernández-Breis Frutos-Morales, Gil, Castellanos-Nieves, Valencia-García, García-Sánchez, and Sánchez-Vera (2011) incorporate semantic interpretations of the mapping between domain ontology and ontology that supports the overall structure of the assessments. Jia, Wang, Ran, Yang, Liao, and Chiu (2011) focus on the particular issues of learning in work environments, learning that is constructed on practical tasks and work situations. The developed ontology assists in facilitating and directing learning activities of employees to achieve individual and organizational needs. Chen, Lee, and Tsai (2011) propose the use of a personalized learning that incorporates an evaluation scheme that generates knowledge questions and correct answers using an ontology in order to assess the level of competence of vendors with respect to knowledge about mobile phones. In her master thesis, Romero and Leone (2007a, 2007b) defined an ontology that emphasizes the design and management of learning material to evaluate the e-learning process in distributed environments. However, this ontology is strongly associated with a specific e-learning platform and may not be reused. Alsubait, Parsia, and Sattler (2012) show the design of an algorithm for generating questions automatically. This algorithm considers only multiple-choice questions, without specifying ontologies on which it is based. Radenković, Krdžavac, and Devedžić (2009), used a descriptive logic reasoner for assessment. This reasoner is focused on the answers and solutions proposed by the students and the intelligent analysis of them.

Scoreanzi and Romero (2014) conducted a comprehensive bibliographic review on the use of semantic web technologies on the e-assessment domain. It proposes a conceptual model, which gives a started point for stakeholders on e-learning projects to know the progress on this subject.

The main objective of this paper is to show the AONET ontology network and discuss in detail the ontologies that conceptualize the assessment domain. To this end, this paper is organized as follows. First the main concepts around semantic web technologies are introduced. Following, the AONET ontology network is presented, and the ontologies of the assessment domain are discussed in depth. Next, the evaluation of the ontologies is presented. In the next section, the use and limitations of the ontologies are discussed. Finally, Section 5 summarizes and concludes this work.

**Background technologies**

The aim of this section is to present the main concepts around semantic web technologies necessary to understand the AONET ontology network.

The concept of “ontology” seems to enjoy as many definitions as attempts to define it. In the course of this paper, we adopt the definition given by Maedche (2002): “An ontology is a 6-tuple consisting of concepts, relations, hierarchy, a function that relates concepts non-taxonomically, a set of axioms, and a set of rules.”

Formally: \( \text{O:} = \{ C, R, H, \text{rel}, A, \text{DR} \} \) where:

- Two disjointed sets, \( C \) (concepts that represents classes) and \( R \) (binary relations among concepts).
- \( \text{H} \subseteq C \times C \) is the concept hierarchy or taxonomy. So, \( H(C1, C2) \) means \( C2 \) subsumes \( C1 \) if and only if \( C1 \) is a subclass of \( C2 \) or if every instance (individual) of \( C1 \) is also an instance of \( C2 \) (Guarino & Welty, 2004).
- \( \text{rel} \rightarrow C \times C \) relates the concepts non-taxonomically.
- \( A \) : set of axioms that are logical sentences which are always true, and express the properties of model paradigm, expressed in an appropriate logical language.
- \( \text{DR} \) a set of derivation rules that model rules about the domain of discourse, expressed in an appropriate logical language. Every \( \text{DR} \) is a Horn-like clause with the structure of \( \forall \phi(x_i) \land \ldots \land p_n(x_i) \rightarrow q(y) \)
Ontologies can be lightweight or heavyweight (Gómez-Pérez, Fernandez López, & Corcho, 2004). The former describes concepts and relationships that hold among them. The latter add axioms and derivation rules to the former. Thus, every heavyweight ontology has a lightweight version.

The main component of a heavyweight ontology is the set of derivation rules. The ontology web language (OWL) (W3C, 2012) is the standard for implementing an ontology. We must combine OWL with other representation formalism in order to generate rules. One of the integration approaches is the semantic web rule language (SWRL), which provides the ability to express Horn-like rules in terms of OWL concepts. With the aim of extracting information from OWL ontologies, a query language is needed. The most powerful language is SQWRL, which is based on the SWRL rule language and uses SWRL’s strong semantic foundation as its formal underpinning (Allocca D’aquino & Motta, 2009).

An ontology network is a set of ontologies related together via a variety of different relationships such as mapping, modularization, version, and dependency. The elements of this set are called networked ontologies (Allocca et al., 2009).

An ontology network differs from a set of interconnected individual ontologies in the relations among ontologies since in an ontology network, the meta-relationships among the networked ontologies are explicitly expressed (Gómez-Pérez, et al., 2004). The DOOR (descriptive ontology of ontology relations) ontology defines general relations between ontologies such as includedIn, equivalentTo, similarTo, and versioning using ontological primitives and rules (O’Connor & Das, 2009). In this paper, the following relationships are used:

- **isTheSchemaFor**
  keeps the link between a model and its metamodel.

- **usesSymbolsOf**
  This relationship happens when an ontology, O, involves individuals belonging to another ontology, Q, in such a way that ontology Q defines some properties that take value in individuals that are classified by classes of ontology O.

- **cross-ontologySubsumption**
  This relationship exists between a class belonging to an ontology and its superclass belonging to another ontology.

- **relatedTo**
  This relationship exists between two classes belonging to different ontologies.

This work was implemented through the NeOn (http://www.neon-project.org) toolkit and complemented by the NeOn methodology, which is a scenario-based methodology that supports the collaborative aspects of ontology development and reuse (Vitturini, Benedetti, & Señas, 2011).

The main advantage of using an ontology network is the conceptualization of a given domain in a modular way. The networked ontology is small enough to be understandable by anybody and its maintenance is easy. In addition, several ontology designers could work on different networked ontologies concurrently.

### The AOnet ontology network development

In order to design the AOnet, the NeOn methodology in conjunction with the methodology proposed by Corcho, Gomez-Pérez & Fernandez-Lopez (2003) was followed. AONet consists of five ontologies that conceptualize three different domains: course topic domain, educational resources, and assessment. Figure 1 shows the AONet ontology network.

The **Educational Resource Specification Ontology** conceptualizes the educational resources used by the educator in the TL process. Some standards emerge to overcome the formalization of educational resources, which are constantly evolving. In most cases, the use of a learning object (LO) definition and its description by LOM (Learning Technology Standards Committee, 2002) is the common denominator. It is possible to optimize the educational resource development process. This ontology is related to Course Domain Specification ontology through the **relatedTo** relationship.
relationship. It identifies the connection between educational resources and concepts belonging to the specific domain. That is to say, an educational resource is developed in order to achieve different concepts, relations and definitions about a domain topic. A LO metadata instance describes relevant characteristics of an educational resource, with the aims of facilitating the search, acquisition, interchange, and evaluation of a resource by educators, learners, and software systems. We add LOnto ontology to the network built (Romero & Godoy, 2010), which conceptualizes the semantic definition of LO based on LOM IEEE 1484.12.1 standard (Learning Technology Standards Committee, 2002). Then, the Educational Resource Specification ontology is related with LOnto through the isTheSchemaFor relationship.

Assessments are part of the educational resources involved in the TL process when the educator wants to evaluate the knowledge and skills acquired by learners. In this context, the ontology network has the Assessment ontology, which is related with Educational Resource Specification ontology through the cross-ontologySubsumption relationship. In the same way, this ontology is related with Educational Domain Specification ontology through the relatedTo relationship. These relations indicate that an assessment is used to evaluate the results of the TL process about the knowledge domain.

There are different instruments to evaluate learning processes, which are modeled by the Assessment Instrument ontology. These instruments are used by educators to create an assessment. For instance, an instrument is a true-or-false question, a conceptual map, an exercise, an essay activity, etc. Using such instruments, the Assessment ontology has the usesSymbolsOf relationship with the Assessment Instrument ontology.

The next sub-sections describe in detail the core ontologies that conceptualize the assessment domain: Assessment Ontology and Assessment Instruments Ontology. In order to develop these ontologies, we followed Scenario 1 from the NEON methodology, since this scenario refers to the development of ontologies from scratch. In this scenario, ontology developers should first specify the requirements that the ontology should fulfill, by means of the ontology requirements specification activity. Then, the ontology developers assigned to the ontology project should carry out: (1) the ontology conceptualization activity, in which knowledge is organized and structured into meaningful models at the knowledge level; (2) the ontology formalization activity, in which the conceptual model is transformed into a semi-computable model; and (3) the ontology implementation activity, in which a computable model (implemented in an ontology language) is generated.
Assessment ontology

The requirements specification of the assessment ontology

This ontology models the structure of an assessment in a TL process and considers the e-learning domain. From a didactics point of view, assessment is necessary to evaluate the learning process and, thus, is also of relevance to the e-learning situation. However, assessment can be considered difficult within a distance learning environment. The assessment can be considered as an element that provides information to guide and enhance the TL process.

With the aim of defining the requirements of the Assessment Ontology, a set of competency questions (CQs) were defined. CQs consist of a set of questions and answers presented in natural language so that the ontology can answer them correctly (Noy & Hafner, 1997). In order to specify the CQs, we identified intended users and domain experts. In this case, three intended users and two domain experts were involved in the CQs definition. The middle-out technique was followed to define the CQs. That is, first, the intended users and domain experts suggested a set of CQs. From this set, complex CQs were decomposed into simple CQs, and simple CQs were organized into complex CQs. During the overall process, we received input from the domain experts and the end users. They used the following criteria for validating the competency questions:

- Correctness. Domain experts checked the correctness of each competency question, verifying that its formulation and answers were correct.
- Consistent. Domain experts also verified that the competency questions did not have any possible inconsistency.

Following, a set of the CQs that the Assessment Ontology has to answer is shown.

- Given an assessment, who is the author?
- Given an assessment, in which moment does it take effect?
- Given an assessment, who is being assessed?
- Given an assessment, who is the evaluator?
- Given an assessment, which activities are part of it?
- Given an assessment, which rubric has been defined?
- Given an activity, which reactives are part of it?
- What are the formative assessments?
- What are the diagnostic assessments?
- Given an author, which diagnostic assessments did he develop?
- Given a student, which self-assessments did he answer?
- What is the role of an agent involved in a given hetero-assessment?

The assessment ontology conceptualization

An Assessment is composed of one or more activities (exercises that assess a particular domain topic). Each activity is composed of one or more reactives (an item that uses an instrument).

There is a moment in the TL process when the assessment takes effect (Arredondo, 2002). If the assessment is early in the process, it is known as a diagnostic assessment. Teachers can use this kind of assessment to become acquainted the background knowledge of student.

A formative assessment is performed during the TL process. Its goal is to give feedback to learners and educators about how well learners understand specific content.

A summative assessment occurs at the end of the process. Its goal is a judgment: to derive a grade, qualification, or accreditation.

Agents (an educator, learner, author, or management) are involved in the assessment-playing roles. These agents can play different roles such as: assessed or assessor.

Depending on the roles that each agent assumes in a particular assessment, the assessment can be classified into Self-assessment, Hetero-assessment, and Co-assessment.
Self-assessment is the assessment in which a learner evaluates his/her own progress in the learning process. In this case, the learner assumes both roles: assessed and evaluator.

Hetero-assessment is a kind of assessment where the agents involved are educator and learner, and the educator assesses a learner. This is a classical approach and is traditionally presented in the TL process, where educator plays the role of evaluator and the learner plays the role of assessed.

Co-assessment is peer assessment. It can be done between learners or between educators playing both roles. Figure 2 shows the ontology model. It is considered a good practice that the rules and style of lexical encoding for naming elements be homogeneous within the ontology. All concepts were defined in singular, and the name of a class was defined using the upper camel case style, while the name of a relationship was defined using the lower camel case style.

The main concept of the Assessment Ontology is the class Assessment. This is a prerequisite for achieving educational quality. Regarding assessment composition, the isComposedBy relation set an Assessment is composed of Activities, which in turn have one or more Reactives.

![Assessment ontology diagram](image)

The Moment class has three instances, Diagnostic, Summative, and Formative, which represent the classification of an assessment according to the moment in which it takes effect. Three Assessment subclasses are defined: SelfAssessment, HeteroAssessment, and CoAssessment. The Agent class represents the actors involved in the assessment process. Four subclasses have also been defined: Author, Learner, Educator, and Management.

The hasEvaluator and hasAssessed relations restrict the agents involved in each type of assessment. Subsequently, hasEvaluator and hasAssessed relations link SelfAssessment with Learner classes. In the same way, the hasEvaluator relates HeteroAssessment with Professor and hasAssessed relates HeteroAssessment with Learner classes. Also, the CoAssessment class has four relations instead of two. One pair of hasEvaluator and hasAssessed relations relates CoAssessment with Educator classes meaning that the assessment takes effect between peer educators. The other
pair of relations relates CoAssessment with Learner classes, meaning that the assessment takes effect between peer learners.

The assessment ontology formalization

Before implementing the ontology, we formalized axioms to restrict the way in which relations can be performed because the definition of relations alone is not enough to define type of assessment according to the agents involved. For instance, CoAssessment can be seen as both HeteroAssessment and SelfAssessment because nothing restricts the definition of the relationships. Table 1 shows the axioms defined to restrict different types of agents involved in SelfAssessment, HeteroAssessment, and CoAssessment. These axioms have been written in first-order logic.

Table 1. Axioms to restrict agents involve in assessment

<table>
<thead>
<tr>
<th>Description</th>
<th>First-order logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A co-assessment takes effect between peer educators or between peer learners.</td>
<td>$\exists_y (\forall x,y,z (\text{CoAssessment}(x) \land \text{hasEvaluator}(x,y) \land \text{hasAssessed}(x,z) \Rightarrow (\text{Educator}(y) \land \text{Educator}(z)) \lor (\text{Learner}(y) \land \text{Learner}(z))))$</td>
</tr>
<tr>
<td>A hetero-assessment has an educator as the evaluator and a learner assessed.</td>
<td>$\exists_y (\forall x,y,z (\text{HeteroAssessment}(x) \land \text{hasEvaluator}(x,y) \land \text{hasAssessed}(x,z) \Rightarrow (\text{Educator}(y) \land \text{Learner}(z))))$</td>
</tr>
<tr>
<td>A self-assessment has a learner which is both evaluator and assessed.</td>
<td>$\exists_y (\forall x,y (\text{SelfAssessment}(x) \land \text{hasEvaluator}(x,y) \land \text{hasAssessed}(x,y) \Rightarrow \text{Learner}(y))$</td>
</tr>
<tr>
<td>An assessment always has an author.</td>
<td>$\exists y (\forall x \text{Assessment}(x) \Rightarrow \exists y (\text{hasAgent}(x,y) \land \text{Author}(y)))$</td>
</tr>
</tbody>
</table>

The assessment ontology implementation

Assessment ontology has been implemented in Protégé ontology editor. Figure 3(a) shows the ontology hierarchy and Figure 3(b) shows the hasAssessed relation properties.

![Figure 3. Assessment ontology implementation](image-url)
Assessment instruments ontology

The requirements specification of the assessment instruments ontology

The Assessment Instrument ontology models different instruments that could be used for implementing an assessment. An assessment instrument is the physical support that is used to collect the information about the expected learning of students.

The Assessment Instrument ontology has to be able to answer the following competency questions:

- Given an assessment, which activities use multiple-choice instrument?
- Given an assessment, which activities use simple-choice instrument?
- Given an assessment, which activities use completion instrument?
- Given a simple-choice instrument, which is the true option?
- Given a simple-choice instrument, which are the distractors?
- Which is the answer of a given correspondence instrument?
- Which are the single instruments of a portfolio?
- What is the organization of a portfolio?
- Which instruments can be used for an assessment?
- What are the types of essay?
- What is the scope of a restricted essay?
- How is a restricted essay organized?
- Which citations have a restricted essay?
- Which are the mechanics of a restricted essay?

The assessment ontology instruments conceptualization

Class Instrument is the main concept in this ontology. It is related to the Reactive concept from Assessment ontology through the uses relation, which in term is an instance of usesSymbolsOf meta-relationship. There are two types of instruments: FormalInstrument and SemiformalInstrument, representing formal and semiformal techniques, respectively. For semiformalInstrument, we have considered two types: SimpleInstrument, such as Exercise, ConceptualMap, and Essay, and CompositeInstrument, such as Portfolio, which consists of a collection of SimpleInstrument elements that help record learning processes and students’ progress. Figure 4 shows the relationships among concepts previously defined.

![Figure 4. Assessment Instrument Ontology](image-url)
As FormalInstrument, we consider two classifications: ObjectiveActivity, where students have to identify the correct answer and EssayActivity, where learners have to elaborate the answer in writing. ObjectiveActivity is one of the most used activities by professors because it eliminates the subjectivity in the rating, even when it has an additional complexity to develop it.

ObjectiveActivity has three sub-concepts: Choice, Correspondence, and Completion. Choice has Option associated. The concept Option is specialized into two sub-concepts: Distractor and TrueOption. Distractors are items that are not correct, and TrueOption is the correct item. The concept Choice is specialized as: SimpleChoice (contains only one correct option), and MultipleChoice (which can have more than one correct option). In both cases, Option can only have a true/false answer associated. Finally the concept Answer can be of different types: TrueFalse, Numeric, Text, and Relation. Figure 5 shows the definition of Objective activity as part of Assessment instrument ontology.

According to Diaz and Barriga (2008), there are some pedagogical recommendations when a choice instrument is used. If educators follow these guides, we can say that the instrument is valid in a pedagogical sense. In this work, we used these recommendations to define rules to express the restrictions in the generation of valid instruments.

From a pedagogical perspective, there should always be a right option. As well, these types of activities should not include options such as “none of the above” or “all of the above.” In general, items should be relevant to the context being assessed in a clear and simple way and preferably written in the affirmative mode. The distractors should appear as attractive as possible to the uninformed student. These pedagogical recommendations can be defined as:

- A simple-choice instrument should have at least four options.
- A simple-choice instrument should have only one true option.
- A multiple-choice instrument should have more than one true option.
- A multiple-choice instrument should have at least four options.
- A multiple-choice instrument cannot have options such as “all of the above” or “none of the above.”

<table>
<thead>
<tr>
<th>Table 2. Pedagogical rules for simple and multiple choice in first-order logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
</tr>
<tr>
<td>Simple choice</td>
</tr>
<tr>
<td>1. A simple-choice instrument should have at least four options.</td>
</tr>
<tr>
<td>2. A simple-choice instrument should have only one true option.</td>
</tr>
</tbody>
</table>
Multiple choice

3. A multiple-choice instrument should have more than one true option.

4. A multiple-choice instrument should have more than four options.

5. A multiple-choice instrument cannot include an option such as “all of above” or “none of above.”

\[ \forall x, y, z (\text{MultipleChoices}(x) \land \text{hasOption}(x, y) \land \text{hasOption}(x, z) \land \text{TrueOptions}(y) \land \text{TrueOptions}(z) \Rightarrow y \neq z) \]

\[ \forall x, y, z, w, r (\text{MultipleChoices}(x) \land \text{hasOption}(x, y) \land \text{hasOption}(x, z) \land \text{hasOption}(x, w) \land \text{hasOption}(x, r)) \Rightarrow (y \neq z \neq w \neq r \land z \neq w \neq r \land w \neq r) \]

\[ \forall x, y, z, w (\text{MultipleChoices}(x) \land \text{hasOption}(x, y) \land \text{hasAttribute}(y, z) \land \text{value}(z, w)) \Rightarrow (w \neq \text{“all of the above”} \land w \neq \text{“none of the above”}) \]

Table 2 shows the pedagogical rules that have been taken into account. The first column describes the rule in a colloquial language. The second column shows the first-order logic description of such rules. For simplicity, in the second column we use a reified concept such as SimpleChoices and MultipleChoices, depicting the sets of simple-choice and multiple-choice elements, respectively.

The EssayActivity is divided into two sub-concepts: RestrictedEssay and UnrestrictedEssay. Restricted essay is a written learning exercise that has a predefined topic and scope (Attali, Lewis & Steier, 2013; Birkenhauer, 2008). All students completing a restricted essay write about the same thing, in the same way, and in accordance with specific instructions and parameters (Landauer, Laham & Foltz, 2003). A rubric is provided to learners so that they can adapt their writing to the restrictions required for the essay (Cooper & Gargan, 2009). The essay must have relevant content, logical organization, and sound written mechanics (i.e., correct spelling, grammar, and punctuation) (Shermis, Mzumara, Olson, & Harrington, 2001). Assessment is conducted in accordance with the rubric (Cooper & Gargan, 2009; Read, Francis, & Robson, 2005). Citation of sources in a specific format (e.g., APA) may or may not be required (Birkenhauer, 2008). Figure 6 shows the description of RestrictedEssay as part of the Assessment Instrument ontology.

In this context, integrity axioms were defined in order to set the condition that a restricted essay must meet. Table 3 shows these axioms.
Table 3. Integrity axioms for RestrictedEssay

<table>
<thead>
<tr>
<th>Description</th>
<th>First-Order Logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A restricted essay must have a topic.</td>
<td>$\mathcal{I} \models (\forall x \text{RestrictedEssay}(x) \Rightarrow (\exists y \text{Topic}(y) \land \text{hasTopic}(x,y)))$</td>
</tr>
<tr>
<td>A restricted essay must have a scopes.</td>
<td>$\mathcal{I} \models (\forall x \text{RestrictedEssay}(x) \Rightarrow (\exists y \text{Scope}(y) \land \text{hasScope}(x,y)))$</td>
</tr>
<tr>
<td>A restricted essay must have mechanics.</td>
<td>$\mathcal{I} \models (\forall x \text{RestrictedEssay}(x) \Rightarrow (\exists y \text{Mechanic}(y) \land \text{hasMechanic}(x,y)))$</td>
</tr>
<tr>
<td>A restricted essay is disjoint with unrestricted essay.</td>
<td>$\mathcal{I} \models (\forall x, y (\text{RestrictedEssay}(x) \land \text{UnrestrictedEssay}(y) \Rightarrow x \neq y)$</td>
</tr>
<tr>
<td>A restricted essay must have organization.</td>
<td>$\mathcal{I} \models (\forall x \text{RestrictedEssay}(x) \Rightarrow (\exists y \text{Organization}(y) \land \text{hasOrganization}(x,y)))$</td>
</tr>
<tr>
<td>A restricted essay must have a content.</td>
<td>$\mathcal{I} \models (\forall x \text{RestrictedEssay}(x) \Rightarrow (\exists y \text{Content}(y) \land \text{hasContent}(x,y)))$</td>
</tr>
<tr>
<td>A restricted essay must have citations.</td>
<td>$\mathcal{I} \models (\forall x \text{RestrictedEssay}(x) \Rightarrow (\exists y \text{Citation}(y) \land \text{hasCitation}(x,y)))$</td>
</tr>
</tbody>
</table>

An unrestricted essay is a written learning activity that may or may not have a predefined topic and scope (Birkenhauer, 2008; McNamara, Crossley, & McCarthy, 2010). Students completing an unrestricted essay may write about similar topics or different ones; however, the objective of the assigned essay must be the same for all learners (Read et al., 2005). Student essays must conform to instructions and parameters; however, these may be intentionally ambiguous or loosely defined in order to encourage discovery or creativity on the part of the learner (Landauer et al., 2003; McNamara et al., 2010). A rubric may be provided to learners so that they can more clearly understand expectations for the essay (Cooper & Gargan, 2009; Shermis, Mzumara, Olson, & Harrington, 2001). The essay must have relevant content, logical organization, and sound written mechanics (i.e., correct spelling, grammar, and punctuation) (Shermis, Shneyderman, & Attali, 2008). Assessment is conducted in accordance with articulated expectations, and, if a rubric is provided, assessment conforms to the rubric (Attali et al., 2013; Shermis et al., 2001). Citation of sources in a specific format (e.g., APA) may or may not be required (Birkenhauer, 2008).

A portfolio (see Figure 7) is a collection of a learner’s evidence of learning in a course or curriculum. A portfolio may contain artifacts created by the learner, solutions to problems solved by the learner, reflections on artifacts or solutions, written descriptions of artifacts or solutions, and reflections on previous feedback from peers or instructors (Chang, Liang, & Chen, 2013; Van der Schaaf, Baartman, & Prins, 2012). A portfolio must have content and organization (Vance et al., 2013). A portfolio’s definition, content, and organization may be proscribed or learner designed (Van der Schaaf et al., 2012; Vance et al., 2013). A portfolio is assessed in accordance with the requirements set forth by the instructor (Vance et al., 2013).

![Figure 7. Semiformal instrument](image-url)
The implementation of the assessment instruments ontology

We implemented the Assessment Instrument ontology in the Protégé ontology editor within the Assessment Ontology. Also, we implemented rules showed in Table 2 as derivation rules by using the SWRL and SQWRL languages. Beginning with simple choices, we implemented rules 1 and 2 as follows:

1. 
   \[ \text{SimpleChoice}(?sc) \ni \text{Option}(?o) \ni \text{hasOption}(?sc, ?o) \ni \text{sqwrl:makeSet}(?os, ?o) \ni \text{sqwrl:groupBy}(?os, ?sc) \ni \text{sqwrl:size}(?t, ?os) \ni \text{sqwrl:greaterThanOrEqual}(?t, 4) \ni \text{optionQuantityValid}(?sc) \ni \text{answerQuantityValid}(?sc) \]

2. 
   \[ \text{SimpleChoice}(?sc) \ni \text{distractor}(?d) \ni \text{sqwrl:makeSet}(?s1, ?d) \ni \text{sqwrl:groupBy}(?s1, ?sc) \ni \text{sqwrl:size}(?t, ?s1) \ni \text{sqwrl:equal}(?t, 1) \ni \text{answerQuantityValid}(?sc) \]

Regarding multiple choice, we have three restrictions 3, 4, and 5 in Table 2 that are represented with sentences 3, 4, 5, and 6, respectively. Note that the restriction 5 was represented with two sentences, 5 and 6, for simplicity.

3. 
   \[ \text{MultipleChoice}(?mc) \ni \text{distractor}(?d) \ni \text{sqwrl:makeSet}(?s1, ?d) \ni \text{sqwrl:groupBy}(?s1, ?mc) \ni \text{sqwrl:size}(?t, ?s1) \ni \text{sqwrl:greaterThan}(?t, 1) \ni \text{answerQuantityValid}(?mc) \ni \text{optionQuantityValid}(?mc) \]

4. 
   \[ \text{MultipleChoice}(?mc) \ni \text{Option}(?o) \ni \text{hasOption}(?mc, ?o) \ni \text{sqwrl:makeSet}(?os, ?o) \ni \text{sqwrl:groupBy}(?os, ?mc) \ni \text{sqwrl:size}(?t, ?os) \ni \text{sqwrl:greaterThan}(?t, 4) \ni \text{optionQuantityValid}(?mc) \]

5. 
   \[ \text{MultipleChoice}(?mc) \ni \text{Option}(?o) \ni \text{hasOption}(?mc, ?o) \ni \text{label}(?o, ?l) \ni \text{sqwrl:normalizeSpace}(?n, ?l) \ni \text{sqwrl:stringEqualIgnoreCase}(?n, \text{“all of the above”}) \ni \text{withoutAll}(?mc) \]

6. 
   \[ \text{MultipleChoice}(?mc) \ni \text{Option}(?o) \ni \text{hasOption}(?mc, ?o) \ni \text{label}(?o, ?l) \ni \text{sqwrl:normalizeSpace}(?n, ?l) \ni \text{sqwrl:stringEqualIgnoreCase}(?n, \text{“none of the above”}) \ni \text{withoutNon}(?mc) \]

Finally, if a simple choice meets the restriction (1) and (2) we can say that this simple choice is valid. This statement is represented with the following rule:

7. 
   \[ \text{SimpleChoice}(?sc) \ni \text{optionQuantityValid}(?sc) \ni \text{answerQuantityValid}(?sc) \ni \text{validSimpleChoice}(?sc) \]

In the same way, if a multiple choice meets the restriction (3), (4), (5), and (6) is a valid multiple choices:

8. 
   \[ \text{multipleChoice}(?mc) \ni \text{withoutAll}(?mc) \ni \text{withoutNon}(?mc) \ni \text{optionQuantityValid}(?mc) \ni \text{answerQuantityValid}(?mc) \ni \text{validMultipleChoice}(?mc) \]

All the axioms were implemented as restrictions. Figure 8 shows the implementation of the axioms defined in Table 3.

Figure 8. Axioms as restrictions in OWL2 language
Evaluation of the assessment and assessment instruments ontologies

To carry out the evaluation of the Assessment and Assessment Instrument ontologies, we considered two aspects. On the one hand, the model is an ontology so it is necessary to evaluate the ontological characteristics to assess the quality of the ontology. On the other hand, ontologies have to be evaluated against a frame of reference to assess the correctness of them (Suárez de Figueroa Baonza, 2010).

Evaluating the quality of the ontologies

The evaluation of an ontology quality is a process to be carried out during the whole ontology building process. The goal of evaluation is to detect errors in modeling characteristics of the assessment domain; i.e., what it does not define or what it defines incorrectly.

With the aim of evaluating the Assessment and Assessment Instrument ontologies as an ontology itself, the inference engine Pellet was used because it is a consolidate tool that is mature in the task of verifying inconsistencies in ontologies. Besides, Pellet can be installed as a plugin in Protégé, is totally compatible with the OWL2 language. Using Pellet, we verified the consistence of the ontologies. Those inconsistencies could be related to the class disposition (classes in the same hierarchy and disjoint classes), the relationship among classes (range and domain), the type of attribute, or the application of rules in the ontology.

In addition, a tool, called OOPS! (OntOlogy Pitfall Scanner, http://oeg-lia3.dia.fi.upm.es/oops) was used to verify a catalogue of common pitfalls that could lead to modeling errors proposed by Poveda Villalon, Suárez-Figueroa & Goméz-Pérez (2010).

Evaluating the correctness of the ontologies

For evaluating individual ontologies, the most common evaluation approaches are (i) to compare the ontology to a gold standard ontology; (ii) to compare the ontology with a source of data about the domain to be covered (e.g., a set of documents); (iii) to evaluate the ontology verifying if it answers the defined competency questions; (iv) to have the ontology evaluated by human experts who assess how the ontology meets the requirements; and (v) to use the ontology in an application and evaluate the results (Suárez de Figueroa Baonza, 2010). To evaluate the correctness of the Assessment and Assessment ontologies, we used the last three approaches.

During the building process of the Assessment and Assessment Instrument ontologies, we evaluated them using the competency questions. In order to do that the evaluation, we take as example, the final exam of Artificial Intelligence course (Figure 9) (www.ai-class). This is a summative assessment. The first activity is about a search domain topic and has two reactives. The latter is about the machine learning domain topic and has one reactive.

To populate the ontologies, we generated an instance of HeteroAssessment as shown in Figure 10. Also, the ExamIntroductionToAI instance has two activities as components: MachingLearningActivity and SearchActivity. The
latter has two reactives associated: SearchItem1 and SearchItem2. The instance of HeteroAssessment has a moment associated as well.

Figure 10. Population of the Assessment ontology

Following, Figure 11 shows the agents associated with the ExamIntroductionToAI instance. It is a HeteroAssessment, and it has associated two agents: first is an instance of Learner, which is associated through the link hasAssessed, and second is an instance of Educator, associated through the link hasEvaluator.

Figure 11. Agents join in the AI assessment

Next, the reactive searchItem1 uses a completion instrument whose answer is 81, an instance of Numeric (Figure 12).
The competency questions defined for these ontologies were implemented in SPARQL. An example of this implementation is shown in Table 4.

<table>
<thead>
<tr>
<th>Competency question</th>
<th>SPARQL query</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the formative assessments?</td>
<td>PREFIX asse:<a href="http://utn.frsf.edu.ar/assessment/">http://utn.frsf.edu.ar/assessment/</a>\nPREFIX ins:<a href="http://utn.frsf.edu.ar/instrument">http://utn.frsf.edu.ar/instrument</a>\nSELECT ?assessment\nWHERE \n{?assessment asse:moment ?moment FILTER{regex(?moment, &quot;formative&quot;) } } }</td>
</tr>
</tbody>
</table>

We evaluated the CQs answers in conjunction with the experts in pedagogical aspects and the professor of the artificial intelligence course at the Universidad Tecnológica Nacional, Facultad Regional Santa Fe, who is an intended user.

After all the competency questions were evaluated, the Assessment and Assessment Instruments ontologies were used as informational source to develop a software tool called OFGA, which allows educators to generate an assessment and check its validity according to the pedagogical rules.

The tool was developed in Java language, using MySQL as DB engine and Apache JENA reasoner. Figure 13 shows the software tool architecture. The DAOs component is the access data layer and the ONTOLOGIES component gives support to the ontologies’ access. Both components provide services to the SERVICES component. View component is the user interface, and MODEL component is the conceptual model used by components in order to communicate with each other.
Figure 13. Software tool architecture

In this way, the ontology creation and manipulation is processed in main memory using only the necessary components in a way to avoid overload. The JENA inference engine allows the tool to derive knowledge using SWRL rules. The tool generates a RDF graph from ontology, which is used by the inference engine to make a reasoning. Then the persistent data are located in a database for future uses.

The main features of OFGA allow the user to generate a new assessment, add new activities, select instruments, validate assessment based on a set of selected rules, add new validations rules, and generate a pdf assessment document, among others. Figure 14(a) shows the assessment edition snapshot. Through this tool, the author can add activities to and erase them from the assessment. Figure 14(b) shows the assessment validation based on a set of selected rules. In this case, there is a set of rules defined, and the author can select which rules to apply to validate the assessment generated. Also, authors can write their own rules and decide whether an assessment meet these rules.

The OFGA was used to validate the Assessment and Assessment Instrument from the intended user’s perspective. While developing these ontologies, we carried out interviews with different experts in pedagogical aspects. To evaluate the ontologies with regard to usability, we invited five educators to use the software to model the evaluation
that was used in their regular courses at the UTN, Facultad Regional Santa Fe, and Universidad Nacional del Litoral. We adopted the following criteria to select participants: all participants belonged to different disciplines; none of the participants had pedagogical training; and all participants had more than 10 years of experience teaching different courses. All participants evaluated the usefulness of the tool, and all of them discovered that most of their assessments did not follow the main pedagogical recommendations. In addition, from this experience we concluded that the Instrument Assessment ontology models most of the instruments types that are commonly used. However, more rules are necessary to evaluate the essay instrument.

Conclusion and future work

This work has shown progress in defining an ontology network whose purpose is to conceptualize the assessment domain in a TL process. The modularization that this network provides allows us to concentrate on a particular domain and incrementally build a more general model relating to different ontologies.

We presented concepts related to the assessment domain. Mainly, this work focused on describing the Assessment ontology that models the different classification and types of assessment in an educational context, as well as the different instruments used in assessment development.

Through the ontology network, it is possible to add a new ontology and relate it to an existing one. This work has presented the meta-relations found between three domain ontologies: topic domain, educational resource, and assessment domain ontologies. This development allows the appropriate description of resources to enhance their location and retrieval.

We present the integrity axiom and derivation rules (static and dynamic conditions) to meet two objectives: restrict the way in which an assessment can be developed and introduce pedagogical criteria. In the approach presented in this paper, the pedagogy and the technology are closely linked. In this sense, the ontologies are useful for assisting educators who do not have strong knowledge of pedagogical aspects.

Finally, we discussed an example of the ontology network population using an artificial intelligence assessment and showed snapshots of a tool to generate assessments using these ontologies. In the future, we intend to acquire additional validation assessments for a broad evaluation and refinement of the ontology.

We presented a tool that supports the development of assessment, and are working on improving the ontology network by adding new concepts and relations and new integrity and derivation rules. Additionally, we are developing tests using different types of assessments provided from different knowledge domains.

Acknowledgements

The authors are grateful to Universidad Tecnológica Nacional (UTN), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional del Litoral (UNL), and the Fullbright Program for their financial support.

References


Effects of Annotations and Homework on Learning Achievement: An Empirical Study of Scratch Programming Pedagogy

Addison Y. S. Su12*, Chester S. J. Huang2, Stephen J. H. Yang2, T. J. Ding3 and Y. Z. Hsieh4

1Research Center for Advanced Science and Technology, National Central University, Taiwan // 2Department of Computer Science and Information Engineering, National Central University, Taiwan // 3Department of Materials and Energy Engineering, Ming Dao University, Taiwan // 4Department of Management and Information Technology, Southern Taiwan University of Science and Technology, Taiwan // ncuaddison@gmail.com // shinjia.huang@gmail.com // jhyang@csie.ncu.edu.tw // tjding@mdu.edu.tw // yizeng.hsieh@gmail.com

*Corresponding author

(Submitted July 20, 2014; Revised November 25, 2014; Accepted February 1, 2015)

ABSTRACT

In Taiwan elementary schools, Scratch programming has been taught for more than four years. Previous studies have shown that personal annotations is a useful learning method that improve learning performance. An annotation-based Scratch programming (ASP) system provides for the creation, share, and review of annotations and homework solutions in the interface of Scratch programming. In addition, we combine the ASP system with the problem solving-based teaching approach in Scratch programming pedagogy, which boosts cognition development and enhances learning achievements. This study is aimed at exploring the effects of annotations and homework on learning achievement. A quasi-experimental method was used with elementary school students in a Scratch programming course over a 4-month period. The experimental results revealed that students’ thoughts and solutions in solving homework assignments have a significant influence on learning achievement. We further investigated that only making annotations in solving homework activities, among all other variables (the quantity of annotations, the quantity of one’s own annotations reviewed, the quantity of peers’ annotations reviewed, the quantity of one’s own homework solutions reviewed, and the quantity of peers’ homework solutions reviewed), can significantly predict learning achievements.

Keywords

Annotation, Scratch programming, Learning performance, Homework

Introduction

Students engage in an especially useful practice by making annotations to learning materials. The positive effects of making and sharing annotations on learning Scratch programming have been presented in the study of Su, Yang, Hwang, Huang and Tern (2014). The use of making annotations has a positive influence on learning achievement and it stimulates students’ motivation to learn (Su, Yang, Hwang, & Zhang, 2010). Hsiao and Brusilovsky (2010) thought that annotations could support community feedback in the process of learning Lego. Garner (2009) indicated that Scratch programming is akin to Lego, which is ideal for primary school students.

We developed an annotation-based Scratch programming system, ASP, which allowed students to highlight important points and key concepts. Furthermore, it supports the storage of making, sharing, and reviewing annotations, so students can subsequently find their previous ideas and thoughts. The mechanism provides opportunities for students to critically recall past memories and to think about how to use them. Making and sharing annotations can expand students’ ideas and thoughts, which may lead to better solutions to solving problems. However, students face various problems in learning Scratch, and invariably students try to find particular ways to solve homework problems. To learn problem-solving-based learning methods is a study in learning itself. The incorporation of ASP, within the context of problem-solving-based teaching, enables students to become aware of and to determine their problem-solving skills and Scratch learning needs (Su, Yang, Hwang, & Tern, 2014). The innovative approach also allows students to be able to enhance their learning potential in the area of Scratch programming. All the issues described above have been explored in this paper by conducting an empirical study of the Scratch programming class using ASP tool-supported problem-solving-based teaching approach.
Literature review

Annotation systems supported by Scratch programming pedagogy

The objective of Scratch programming pedagogy is to encourage students to think about their own practical experience. Garner (2009) and believe that Scratch programming should have a “low floor” (easy to get started) and a “high ceiling” (opportunities to create increasingly complex projects over time). Thus, we conducted an empirical study of Scratch programming pedagogy at a Taiwanese elementary school, and students could easily understand the Chinese interface of Scratch programming when first introduced to Logo programming.

A variety of annotation systems have been developed for helping teachers and students to teach and learn digital learning materials. Microsoft has published OneNote that is a personalizable annotation tool. Teachers and students can use a digital pen, a keyboard, or a multimedia recording to input annotations. However, the system annotations can only annotate Microsoft documents and be shared through e-mails. Yang, Chen, and Shao (2004) proposed a personalized annotation management system, and they have demonstrated that personalized annotation enhances knowledge sharing and learning in students’ online learning activity. Hwang, Wang, and Sharples (2007) developed a multimedia annotation system that enables students to annotate multimedia objects as well as create multimedia content, such as pictures, audio and video as annotations. Yang, Zhang, Su, and Tsai (2011) proposed a collaborative multimedia annotation system that allows users of audio and video to make and share personal and group annotations that are synchronized with the multimedia. The system permits sharing of personal knowledge and collaboration among learners; annotations can be accessed by other users remotely through the Internet.

Following the previous studies, we found that existing annotation systems do not support the Chinese interface of Scratch programming, and users cannot make, review, and share annotations directly inside the Scratch programming editing environment. In this study, we modified existing annotation systems to provide for creating, reviewing, and sharing annotations in the interface of Scratch programming. The annotation system supported Scratch programming pedagogy is defined as the process wherein students make, review, and share annotations so they can subsequently review their previous lecture notes and homework solutions. This mechanism provides opportunities for students to critically re-evaluate and possibly improve their thoughts and solutions. Finally, we aimed to use the ASP system supported problem solving-based teaching approach in learning Scratch programming.

Influence of annotations and homework on learning achievement

Slotte and Lonka (1999) indicated that students who created annotations generally attained greater levels of academic achievements than those who did not; students understood the learning material deeper and they performed better on a test. Yeh and Lo (2009) conducted an experiment for giving feedback on second language writing of college freshmen by creating, sharing and reviewing annotations in homework activities. Moreover, the result showed that students who were arranged in an annotation tool group had significantly better error recognition learning performance than those who were arranged in a traditional paper-based annotation group.

As for annotations reviewed by other peers, Slotte and Lonka (1999) emphasized that reviewing annotations have the advantage of allowing for informal knowledge sharing related to a concept. Nokelainen, Miettinen, Kurhila, Floréen, and Tirri (2005) thought that reviewing annotations allows learners to collaborate by sharing their thoughts about the learning material with other students accessing the same content. Reviewing homework solutions in this study is defined as the process when students reviewed peers’ homework. By reviewing annotations and homework of other peers, students may benefit from peer learning via critical peer feedback and reflection.

As for homework, Power, Dombrowski, Watkins, Mautone, and Eagle (2007) reported that homework in elementary school may be completed in school and at home after school. Power et al. (2007) argued that homework has positive effect because homework has an immediate effect on the retention and understanding of the material it covers. Based on previous studies, homework is not only helpful for enhancing individual learning, but it is a useful way to facilitate peer learning through sharing annotation progresses.

In this study, we applied the developed ASP system to support students with making, reviewing, and sharing annotations and solving homework during learning Scratch programming activities. This study aimed to explore the
effects of creating annotations and homework solutions, reviewing own and peers’ annotations, reviewing own and peers’ homework solutions on learning achievement.

Problem-solving method supported Scratch programming pedagogy

Papert (1980) indicated that students need to possess a suitable problem-solving method, in which people try to make sense of a problematic situation using previous knowledge and practical experience by attempting to acquire relevant information and solutions about that situation until they can resolve the ambiguity. The problem-solving-based instructional method plays an important role in helping students learn better; it helps understand how the task is performed (Hung, 2008). The empirical studies of existing problem-solving-based instructional methods include at least four distinct steps: (1) understanding and defining the problem; (2) planning solutions and collecting necessary information; (3) designing and implementing the solution; and (4) verifying and presenting results.

Hung (2008) have contended that existing problem-solving-based instructional methods need to correspond with school children in terms of Scratch programming studies. According to existing problem-solving-based teaching methods, we refined a way to accomplish the objective in four distinct steps. Step 1: Explain the objective of an example. Step 2: Understand the requirement of the example. Step 3: Propose a plan to solve the design of the example. Step 4: Implement the example and debug it. Adherence to these steps strongly promoted students’ problem-solving skills. Hung (2008) thought that an appropriate selection of problem-solving instructional strategies plays an important role in helping younger schoolchildren learn; it helps to understand how the programming task is performed. The appropriate method could be used to help schoolchildren establish their learning objectives and assess their progress toward these objectives.

In this study, we sought to facilitate problem-solving abilities and cognitive developments among students by providing a means to which students could create, share, and review annotations and work on homework. We focused on study cognitive skills that would show when students review their annotations and homework solutions using the ASP-supported problem-solving-based teaching approach.

Perceived satisfaction of using ASP-supported problem-solving-based teaching approach

Perceived usefulness and perceived ease-of-use proposed by Davis (1986) have been widely used to predict user attitudes toward information technology. A learner’s behaviour toward an information system was determined by his/her attitude concerning the perceived usefulness and perceived ease-of-use. Perceived usefulness and perceived ease-of-use were evaluated with the initial version of ASP (Yang et al., 2011; Cheng, Lou, Kuo, & Shih, 2013), and our results showed that students perceived the ASP system to be both useful and easy to use. Perceived usefulness refers to the belief that using ASP will boost one’s performance. Perceived ease-of-use refers to the belief that using ASP will be relatively effortless. In this study, a questionnaire, adapted from Hwang, Wang, and Sharples (2007), was employed to gain an understanding of the students’ perceived satisfaction with using ASP-supported problem-solving-based teaching approach.

Research objectives

The initial version of ASP was a web-based annotation system and it was developed by Su, Yang, Hwang, and Zhang (2010). The subsequent version was modified into a multimedia-based annotation system in the following years (Su et al., 2014). The experiment was conducted with the ASP-supported problem-solving-based teaching approach, and the results show that students perceived the system to be both useful and easy to use. Moreover, the students were satisfied with using ASP-supported problem-solving-based teaching approach in individual and collaborative learning scenarios. The learning achievements of students in the experimental group (using ASP-supported problem-solving-based teaching approach) were significantly better than those of students in the control group (using traditional pen and paper) in the two learning scenarios; our results revealed that students were satisfied with using ASP in two learning scenarios.
Based on the previous studies, we conducted a real study (e.g., one Scratch programming course) at a primary school in order to explore the effects of annotations and homework on learning performance. At the end of this course, the questionnaire was given to evaluate the students’ attitudes -- their perceived satisfaction with using ASP-supported problem-solving-based teaching approach.

**The operation of the Annotation-based Scratch Programming (ASP) system**

To efficiently conduct the ASP-supported problem-solving-based instructional approach, four distinct steps are given to students to understand how their homework is solved. The homework example with the topic “A Sprite Walking Staircase” was uploaded to the ASP and made available to students.

Step 1: Explain the goal of a sprite walking staircase, as shown in Figure 1, the teacher reminds students to recall and employ their past knowledge experience in performing the sprite walking staircase.

![Figure 1. Explaining a sprite walking staircase programming exercise](image1)

**Step 2:** Understand the input and output requirements of a sprite walking staircase, as shown in Figure 2, the teacher made notes in red to provide clues regarding the requirements of the sprite walking staircase.

![Figure 2. Demonstrating a student’s understanding requirements of the sprite walking staircase](image2)
Step 3: Collect necessary information and propose a plan to finish the design of a sprite walking staircase, as shown in Figure 3, students need to understand how loop commands are interpreted by Scratch programming with operations in order to accomplish the repetitive stairs in the example. Furthermore, students indicate the starting point and the repetitive parts of the stairway in light of the demonstrative teaching of the teacher.

Figure 3. Showing students’ proposing a solution plan for designing the sprite walking staircase

Step 4: Design a sprite walking staircase and debug it, as shown in Figure 4, the teacher answers a student’s question in red ink, such as indicating the source of Scratch programming errors in the example.

Figure 4. Students’ writing the sprite walking staircase and debugging it
The ASP provided straightforward access control over annotations that were made and retrieved. For example, a student may be granted read and write privileges to annotations. This access control mechanism provided multiple choices for sharing and reviewing annotations with others; students were able to grant or deny access rights to their peers. In addition, the ASP automatically saved manipulation portfolio data created by all students to our database, which included the students’ comments and thoughts, the students’ answers to each question, the time duration for each operation, and so on. By exploring students’ manipulation of portfolio data, we were able to analyse the detailed students’ learning behaviours while studying Scratch programming.

Methodology

Participants

Thirty-seven students (the average age was 12 years old) from one sixth-grade class of a northern Taiwanese elementary school participated in this experiment. Most of students understood basic computer concepts, but they had no experience in Scratch programming. The teacher, with 12 years of teaching experience, led a course entitled “Scratch Programming.” Chinese was used as a medium of instruction. The course ran three hours per week in a computer classroom for four months, from March to June 2013.

Instruments

Materials

The course covers the Introduction to Scratch programming for primary school students. The teacher prepared two topics of Scratch programming in the following sequence. The first topic included event triggers and variables using the block palette combined with the Sensing and Pen functions in Scratch programming to create a hexagon shape. The second topic included conditionals and loops combined with the Control function in Scratch programming to create staircase and windmill shapes. Due to teaching appointment periods and limitations, the teacher focused on two topics. The teacher designed the instrument used to serve as a Scratch programming achievement test.

Scratch programming achievement test

The test aimed to evaluate the learning achievements of the students at the end of the semester. It consisted of 6 multiple choice-items (10 points each) and 2 problem-solving implementation items (20 points each). The multiple choice-items described the structure, syntax, and implementation concept of Scratch programming, and the correctness of the answer to each problem-solving implementation item was assessed by the teacher. If the item is completely correct, 20 points was given for the correctness criterion. If the item was partially correct, 2 points was deducted per “bug.” At last, if the item could not be executed due to syntax errors, zero points were given. According to Brown (1996) and Su et al. (2014), Kuder Richardson 20 (KR-20) was used to evaluate the internal consistency reliability of the test. The KR-20 coefficient of the test was 0.72, showing adequate internal consistency in evaluating the students’ learning performance.

Questionnaire

The questionnaire mainly surveyed students’ perceived satisfaction of using ASP-supported problem-solving-based teaching approach. During the process of designing a valid questionnaire, applicable items were verified and validated by the two education experts. Some ambiguous or unsuitable items were modified, alternated, or removed accordingly. The questionnaire was composed of response questions with seven items in a 5-point Likert and one open-ended question. According to the reliability evaluation of the questionnaire, the Cronbach α coefficient was 0.83, implying that the questionnaire had a high reliability (Carmines & Zeller, 1979). Thirty-seven students were asked to complete the questionnaire after the post-test.
Procedures

During learning Scratch programming, all young students were asked to make annotations, to solve homework exercises and to review annotations and solutions. School children were volunteered and curious to use the ASP system in the context of learning Scratch programming because it was easy to use, with friendly user interface and many students considered that using annotations were very interesting. The teacher was motivated to use the ASP-supported problem-solving-based teaching approach to teach Scratch programming.

A quasi-experimental method was used with elementary school students in a Scratch programming course over a 4-month period, and the experimental procedures were conducted as follows. Students were asked to study the learning material unit-by-unit and to make annotations on the content. Each unit of learning materials consisted of Scratch programming homework exercises, such as a hexagon shape, a staircase and a windmill shape. Students were asked to complete and solve their homework problems after the teacher taught a unit; they could do it immediately during class. Students were then asked to review their own annotations and homework solutions. Furthermore, the students were asked to share their annotations and homework solutions with other students and review the annotations and homework solutions of their peers.

At the end of the semester, a post-test was administered to evaluate students’ learning performance. After the post-test, the questionnaire was given to understand students’ perceived satisfaction of using ASP-supported problem-solving-based teaching approach.

Data collection and analysis

We collected data to response to the research objectives. The data comprised students’ manipulation portfolio data, post-test scores, and questionnaire data. In class, the students’ manipulation portfolio data was stored in our server database. The students’ manipulation portfolio data provided statistical information for each student on the number of annotations made, the number of one’s own annotations reviewed, the number of peers’ annotations reviewed, the number of homework solutions made, the number of one’s own homework solutions reviewed, and the number of peers’ homework solutions reviewed. The statistical information helped us to understand annotation and homework contributions of each student. Two education experts were involved in content analysis to perform coding for the number of annotations and homework solutions. Content analysis resulted in an inter-rater reliability \( (k = 0.76) \) of more than 0.7, and the analysis differences were resolved by the two experts through discussion. At the end of the semester, a post-test was conducted to obtain students’ learning performance. Furthermore, a questionnaire was given to understand students’ perceived satisfaction with using ASP-supported problem-solving-based teaching approach.

Research variables

Students were taught within the same learning period by the same teacher with the same tools and learning materials provided in this study. Moreover, additional variables were controlled. The number of annotations and homework were explored to show their influence on students’ learning achievement. The stepwise multiple regression analyses were used to predict learning achievement from annotations and homework solutions. The independent variables are: (1) the quantity of annotations, (2) the quantity of one’s own annotations reviewed, (3) the quantity of peers’ annotations reviewed, (4) the quantity of homework solutions, (5) the quantity of one’s own homework solutions reviewed, and (6) the quantity of peers’ homework solutions reviewed. The dependent variable is (7) learning achievement.

The following variables related to the number of annotations and homework solutions were defined, and we explored the relationships of these variables among each other and with students’ learning performance:
(1) The quantity of annotations: The number of annotations made by a student on learning materials.
(2) The quantity of one’s own annotations reviewed: The number of times that a student reviewed their own annotations on learning materials.
(3) The quantity of peers’ annotations reviewed: The number of times that a student reviewed peers’ annotations on learning materials.
The quantity of homework solutions: The number of annotated homework solutions created by a student on homework assignments.

The quantity of one’s own homework solutions reviewed: The number of times that a student reviewed their own homework solutions on homework assignments.

The quantity of peers’ homework solutions reviewed: The number of times that a student reviewed peers’ homework solutions on homework assignments.

Learning achievement: These were measured by the post-test scores earned by the students in the final examination at the end of the course.

Results

Correlation analysis: Annotations, homework, and learning achievement

According to annotations and homework recorded by students’ manipulation portfolio data, we found that the proportion of annotations to learning materials was 0.38, and the proportion of annotations to homework was 0.62. The Pearson’s correlation was utilized to measure the relationships between variables related to annotations, homework, and learning achievement. The results are shown in Table 1 and Table 2 and some findings are discussed below.

Correlation analysis of annotations and learning achievement

The quantity of annotations \((r = 0.552^{**})\) and the quantity of their own annotations reviewed \((r = 0.433^{**})\) were significantly correlated with learning achievement. This finding was explored deeply via content analysis. From content analysis, we found that students who created annotations focused on clarifying important points of the learning materials. Moreover, students usually recollected their own previous knowledge when they experienced difficulties understanding learning material content. Students looked back to their records to recall their memories and to acquire new ideas, which in turn led to a deeper comprehension of learning material content. Meanwhile, an interesting phenomenon found was that a significant correlation existed between the quantity of annotations made and the quantity of the students’ own annotations reviewed \((r = 0.460^{**})\), implying that students would like to look back records because those comments were very meaningful to them. Furthermore, students would review their comments or ideas to find important points or solve problems when studying learning materials.

No significant correlation was found between the quantity of peers’ annotations reviewed and learning achievement \((r = -0.152)\). From content analysis, we found that some school children annotated the learning material overly, such as mark annotations. Furthermore, students reported that they could not easily recall what was understood based on these mark annotations when they returned to peers’ annotations. During the interview, students mentioned that peers’ annotations helped them to understand important hints and concepts of the learning material content.

| Table 1. Correlation statistics of annotations and learning achievement |
|--------------------------|-----------------------------|-----------------------------|-----------------------------|
| Learning achievement     | The quantity of annotations | The quantity of one’s own annotations reviewed | The quantity of peers’ annotations reviewed |
| Learning achievement     | 1                           | 0.552**                     | 0.433**                     | -0.152                      |
| The quantity of annotations | 0.552**                   | 1                           | 0.460**                     | -0.110                      |
| The quantity of one’s own annotations reviewed | 0.433** | 0.460** | 1 | 0.215 |
| The quantity of peers’ annotations reviewed | -0.152 | -0.110 | 0.215 | 1 |

Note. \(^{*} p < .01.\)

Second, the relationships between variables related to homework and learning achievement were analysed using Pearson’s correlation. The results are shown in Table 2 and some findings are discussed below.
Correlation analysis of homework and learning achievement

The quantity of homework solutions was significantly correlated with learning achievement ($r = 0.711^{**}$), implying that the annotations of homework assignments were strongly associated with learning achievement. Students thought that using ASP-supported problem-solving-based teaching approach could boost a deeper understanding of their homework problem. The quantity of students’ own homework solutions reviewed had a significant correlation with learning performance ($r = 0.527^{**}$). Students indicated that reviewing their own annotated homework inspired them to find potential and relevance solutions. The quantity of one’s own homework solutions reviewed had a significant correlation with the quantity of homework solutions ($r = 0.424^{**}$). Through content analysis, it was found that students preferred to review their annotated homework so that solving homework problems became a good learning process. This process could then help students gather relevant knowledge to acquire their potential solutions needed, which belong to a high level of cognition for students’ problem-solving abilities. The quantity of peers’ homework solutions reviewed had no significant correlation with learning achievement ($r = -0.152$). Students reported that they reviewed the homework solutions of their peers when they encountered difficulties in solving homework problems. However, some students reviewed a high frequency of peers’ solving homework problems that mostly consisted of a peripheral review of the ideas. We found that some school children preferred to review and copy the ideas and thoughts of the teacher and classmates when they improved upon their homework solutions. The interviews revealed that students preferred to review peers’ homework solutions to deep classmates’ intelligence of various solving strategies to help solve new homework.

Table 2. Correlation statistics of homework and learning achievement

<table>
<thead>
<tr>
<th></th>
<th>Learning achievement</th>
<th>The quantity of homework solutions</th>
<th>The quantity of one’s own homework solutions reviewed</th>
<th>The quantity of peers’ homework solutions reviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning achievement</td>
<td>1</td>
<td>0.711**</td>
<td>0.527**</td>
<td>-0.152</td>
</tr>
<tr>
<td>The quantity of homework solutions</td>
<td>0.711**</td>
<td>1</td>
<td>0.424**</td>
<td>-0.110</td>
</tr>
<tr>
<td>The quantity of one’s own homework solutions reviewed</td>
<td>0.527**</td>
<td>0.424**</td>
<td>1</td>
<td>0.210</td>
</tr>
<tr>
<td>The quantity of peers’ homework solutions reviewed</td>
<td>-0.152</td>
<td>-0.110</td>
<td>0.210</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. **$p < .01$.

Effects of annotations and homework on learning achievement

Multiple regression analyses were employed to further predict the impact of annotations and homework solutions on learning achievement. The results of the stepwise multiple regression analyses are shown in Table 3 and Table 4. In Table 3, the $R^2$ value ($R^2 = 0.505, F = 35.736, p < .01$) revealed that the model predictability is 50.5%. That is, six independent variables could be used to predict learning achievement: annotations, one’s own annotations reviewed, peers’ annotations reviewed, homework solutions, one’s own homework solutions reviewed, and peers’ homework solutions reviewed.

Table 3. The summary of the multiple regression model

<table>
<thead>
<tr>
<th>Model</th>
<th>$R$</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.711</td>
<td>0.505</td>
<td>0.491</td>
<td>35.736</td>
<td>.000**</td>
</tr>
</tbody>
</table>

Note. **$p < .01$, dependent variable: learning achievement.

The results of stepwise multiple regression analyses are shown in Table 4. It was found that six independent variables related to annotation and homework solutions were in this regression analysis, and only values with significant potentiality were included into the stepwise multiple regression model. Our results showed that the standardized coefficients $\beta$ of “the quantity of homework solutions” was $0.711 (t = 5.978, p < .01)$, which was the highest of all. Only the quantity of homework solutions out of all other variables (the quantity of annotations, the quantity of one’s own annotations reviewed, the quantity of peers’ annotations reviewed, the quantity of one’s own homework solutions reviewed, and the quantity of peers’ homework solutions reviewed) was the significant predictor of learning
achievement. Therefore, the teacher should encourage students to use ASP-supported problem-solving-based teaching approach to search and collect relevant knowledge to finish homework assignments. The teacher should also inspire students to increase the quantity of homework solutions, which can help foster the learning achievement of school children into Scratch programming activities.

Table 4. Stepwise multiple regression model of predicting learning achievement

<table>
<thead>
<tr>
<th>Model</th>
<th>Independent variables</th>
<th>β</th>
<th>B</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>The quantity of homework solutions</td>
<td></td>
<td>0.711</td>
<td>0.079</td>
<td>5.978</td>
<td>.000**</td>
</tr>
</tbody>
</table>

*Note. **p < .01, dependent variable: learning achievement.

Results of questionnaire

Based on the data collected from the questionnaire, we examined some meaningful information and findings from the educational point of view. With respect to students’ perceived satisfaction shown in Table 5 (mean score of all items is higher than 4.0), most students agreed that the ASP-supported problem-solving-based teaching approach could satisfy the needs of learning Scratch programming. The mean score of items 1, 2, 3, 5, 6, and 7 is higher than 4.2, implying that the ASP-supported problem-solving-based teaching approach could help students to easily express their thoughts and to clearly explain their homework problem process. The annotation tool can also support annotation storage and share, so students can subsequently look back their previous comments; it provides opportunities for them to critically find potential solutions and to think about how to improve them.

The average score of item 4 is the lowest in Table 5. It was found that understanding classmates’ homework problems were too difficult and few students could give helpful clues. Furthermore, the teacher might need to provide more help to the few students who did not achieve an understanding of classmates’ homework problems. Based on students’ higher agreements on these items, they expressed higher interests in using ASP-supported problem-solving-based teaching approach for study with their friends after class.

Table 5. Students’ perceived satisfaction of using ASP-supported problem-solving-based teaching approach

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>SA</th>
<th>A</th>
<th>U</th>
<th>D</th>
<th>SD</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I think that the ASP-supported problem-solving-based teaching</td>
<td>13</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.24</td>
</tr>
<tr>
<td></td>
<td>approach can help me to easily express my annotations and to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clearly explain my homework problem process.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>I think that the ASP-supported problem-solving-based teaching</td>
<td>15</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4.35</td>
</tr>
<tr>
<td></td>
<td>approach can give me a suitable problem-solving approach to learn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scratch programming.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>I think that the ASP-supported problem-solving-based teaching</td>
<td>14</td>
<td>19</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td>approach can help me to clear my homework content.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>I think that understanding classmates’ homework problems are easy</td>
<td>12</td>
<td>15</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>4.05</td>
</tr>
<tr>
<td></td>
<td>and I can provide helpful comments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>I think that the annotation tool can provide opportunities for me</td>
<td>13</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4.24</td>
</tr>
<tr>
<td></td>
<td>to critically review my annotations and to think about how to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>improve the quality of my homework solutions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Compared with online homework that did not provide the ASP-</td>
<td>18</td>
<td>13</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>4.49</td>
</tr>
<tr>
<td></td>
<td>supported problem-solving-based teaching approach, homework</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with the innovative approach used in my class did satisfy my</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>functional demands. I had got more happiness in learning Scratch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>programming.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Comparing with online homework that did not provide the ASP-</td>
<td>14</td>
<td>19</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td>supported problem-solving-based teaching approach, homework</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>with the innovative approach used in my class did improve my</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>learning achievement.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this questionnaire, there was one open-ended question to allow students to present their perceived satisfaction of using ASP tool-supported problem-solving-based teaching approach. To evaluate the students’ intentions to accept
the innovative approach, a question was presented: “Do you like to learn Scratch programming via using ASP tool-supported problem-solving-based teaching approach?” The following statements reflect the students’ responses:

“The innovative approach was helpful for learning Scratch programming. When I encountered homework problems, I usually asked the teacher or my classmates to get assistance in solving my problems. Moreover, my annotations made in class gave me important clues to solve my problems.”

“I know I could find some friends online. I took some annotations for my friends when I found something wrong in their homework. To remind my friends take care of mistakes in their homework. If I find someone’s annotations are valuable, I will repeatedly recall his/her comments and update my homework solutions in the other way.”

“I want to view my classmates’ homework solutions to check whether they are better than mine. Also, I think this helps me to find the drawbacks of my homework methods and inspire me to improve the quality of homework solutions.”

“Although I often do not fully understand others’ annotations in Scratch programming studies, I think that the activity will do me good.”

Discussion

In this study, we aimed to investigate the effects of annotations and homework on learning achievement. The findings are summarized as follows.

We investigated the effects of annotations and homework on learning achievement; interpretations of why these results were obtained are presented. The results of stepwise multiple regression analyses showed that only the quantity of homework solutions, out of all other variables (the quantity of annotations, the quantity of one’s own annotations reviewed, the quantity of peers’ annotations reviewed, and the quantity of one’s own homework solutions reviewed, and the quantity of peers’ homework solutions reviewed), is the significant predictor of learning achievement. Although other variables did not significantly predict students’ learning performance, most students believed that they reviewed annotations to recall the content of learning materials and to understand it better. Besides, reviewing peers’ annotations and homework solutions did not correlate with students’ learning performance. This is in accordance with past studies (Hwang et al., 2007; Su, Yang, Hwang, & Zhang, 2010; Tam, 2009) that reviewing peers’ annotations were not found to improve students’ learning performance. The primary reason for this is because peers’ mark annotations are not easy for others to understand; the meaning and understanding of annotations are strongly connected to their creators. Therefore, it is recommended that the teacher should consider using the ASP-supported problem-solving-based teaching approach that can boost problem-solving skills and cognitive development into learning Scratch programming process, such as encouraging school children to take annotations, solving homework, and sharing them.

From the results of the questionnaire, it was found that most students were satisfied with using ASP-supported problem-solving-based teaching approach in the context of learning Scratch programming. The students also had strong desires to use the innovative approach to do their homework, to explain how they solve problems, to recall their acquired learning knowledge, to simulate their new thoughts, and to provide helpful suggestions to others. Students reviewed annotations to reinforce a deeper comprehension of the learning material content and homework problems. However, few students thought that understanding classmates’ homework mistakes were a barrier to giving helpful suggestions. Thus, the teacher would give more help to weaker students by providing useful suggestions and appropriate prizes. Moreover, the innovative approach was crucial for helping students find right peers and right information and for inspiring constructive ideas in order to complete homework.

With respect to the findings related to the open-ended question of this questionnaire, most students mentioned that there were a number of benefits from using ASP tool-supported problem-solving-based teaching approach in learning Scratch programming. Most students perceived that the innovative approach could help them to improve the quality of their homework solutions to explore how to overcome their homework difficulties and mistakes reflected on their learning performance. Furthermore, most students indicated that they liked to view suggestions to recall their
memories and to inspire fresh ideas for resolving their difficulties. Our results revealed that most of the students were positive and considered the innovative approach to be convenient and helpful to learning Scratch programming.

Conclusion

This study developed an annotation-based Scratch programming system, ASP, which was helpful in creating, sharing, and reviewing annotations and homework. We aimed at exploring the effects of annotations and homework on students' learning achievement in an empirical study of Scratch programming pedagogy. Our results are consistent with some earlier studies (Hwang et al., 2008; Nokelainen et al., 2005; Slotte & Lonka, 1999; Su & Yang, 2010; Su, Yang, Hwang, & Zhang, 2010; Su et al., 2014), which indicated that making annotations are significantly correlated with students' learning achievement. Meanwhile, the quantity of one's own annotations reviewed and the quantity of one's own homework solutions reviewed are significantly correlated with students' learning performance. An implication of this is that a deeper understanding of learning materials and homework problems are directly correlated with reviewing products of study, and furthermore, that homework is also helpful in individual learning. The review of annotations and homework solutions written by other peers was found to have no significant influence on students’ learning achievement. Students reported that a primary reason for this phenomenon is the fact that it is hard to understand mark annotations made by others; annotations are usually most meaningful to their creators. We found that some students annotated the learning material overly, such as mark annotations. Furthermore, students reported that they could not easily recall what was understood based on these mark annotations when they returned to peers’ annotations. Only the quantity of homework solutions among all other variables was found to be a significant predictor of students’ learning achievement. Student reported that they preferred to annotate homework problems to deep peers’ intelligence of various solving strategies to help solve new homework. The quantity of peers’ homework solutions reviewed had no significant correlation with learning achievement. However, students said during the interview that reviewing peers’ homework solutions was helpful for them to find some helpful clues. Therefore, further investigation of this phenomenon is necessary.

According to the findings of this questionnaire, most of the students maintained an overall positive perception of using, and were generally satisfied with, the ASP-supported problem-solving-based teaching approach. Most of the students perceived that the innovative approach could stimulate them to figure out more main points and to perceive more incomprehensible problems, which may make them pay more attention during a lecture. Furthermore, doing homework with the innovative approach could increase students' interest, happiness, and achievements in learning Scratch programming. The innovative approach could help students to easily express thoughts and to clearly explain their process for solving homework assignments, in addition to improving interaction among students, annotations, and homework solutions. However, a few students thought that it was hard to understand annotations made by other classmates; they felt that the teacher might need to provide more help and suggestions. Finally, most of the students thought that the innovative approach could satisfy the needs of learning Scratch programming. From the findings of the open-ended question of this questionnaire, most students liked to make and review digital annotations while learning Scratch programming and doing homework assignments. Through the use of the innovative approach, students felt they could more readily find potential resources and information to help solve problems. These findings are consistent with some past studies (Cheng, Lou, Kuo, & Shih, 2013; Hsiao & Brusilovsky, 2010; Su, Yang, Hwang, & Zhang, 2010; Su et al., 2014), which indicated that the innovative approach could enable students to deeply understand the content of the learning material and homework presented in this course.

Based on the findings of the statistical analyses and questionnaire, it is recommended that teachers consider incorporating the innovative methods into their educational environment in order to foster students’ cognitive development in the area of Scratch programming. This can be achieved by encouraging students to make annotations in the homework assignments, followed by subsequent reviews. The creation and review of annotations and homework were found to be beneficial to learning Scratch programming and doing homework assignments; therefore, it might be useful for encouraging students to create and review their annotations more frequently and more deeply. The results of this study may shed light on higher levels of knowledge understanding with the ultimate aim of improving students’ understanding and practice of Scratch programming. The key issue is we provided a tool for sharing annotations and homework but the study focused on benefit of annotations and homework to individual learning only. In future studies, we will adopt collaborative problem-solving strategy to solve group homework, we will evaluate if this strategy is able to improve school children learning performance. Furthermore, we hope to see from our future studies that such annotations will stimulate students to make more annotations and to receive more
benefit from the outputs of their studies. The effects of using annotations within the collaborative problem-solving strategies will be investigated in future study.

Acknowledgements

This work is supported in part by Ministry of Science and Technology, Republic of China, Taiwan under contract numbers MOST 104-2517-S-008-001 and MOST 104-2511-S-008 -011.

References


Learning Style and Task Performance in Synchronous Computer-Mediated Communication: A Case Study of Iranian EFL Learners

Mohsen Hedayati* and Elham Mohammadi Foomani

Faculty of Foreign Languages, University of Tehran, North Amirabad, Tehran, Iran // Mohsen.hedayati@alumni.ut.ac.ir // Foomanielham@ut.ac.ir

*Corresponding author

(Submitted July 31, 2014; Revised December 15, 2014; Accepted March 20, 2015)

ABSTRACT

The study reported here explores whether English as a foreign Language (EFL) learners’ preferred ways of learning (i.e., learning styles) affect their task performance in computer-mediated communication (CMC). As Ellis (2010) points out, while the increasing use of different sorts of technology is witnessed in language learning contexts, it is worth studying the conditions in which the most second language (L2) production would be accomplished. The participants were 40 advanced-level Iranian EFL learners enrolled at a language institute in Tehran. Learners’ individual learning styles were probed by Felder-Soloman (1991) Index of Learning Style (ILS) and they were categorized into 8 groups, within 4 dimensions: Active vs. Reflective, Sensing vs. Intuitive, Visual vs. Verbal, and Sequential vs. Global learners. Then, the participants were given the opinion-gap tasks in 6 consecutive online chat sessions within a 3-week period. The participants’ produced language was analyzed at two levels: vocabulary, and grammar. Independent samples t-test were conducted to check if the differences between the groups were significant. The results reveal that the Reflective learners and Visual Learners produced grammatically more complex and lexically denser sentences than the other groups, which suggests that learners’ learning styles may affect their task performance in synchronous computer-mediated communication.

Keywords

Learning style, Synchronous computer-mediated communication, Task-based language teaching, Opinion-gap task, Felder-Soloman index of learning style

Introduction

Recently, studies on Task-based Language Teaching (TBLT) have lent credence to investigation of TBLT in technology-mediated contexts to find out the technology’s impact on task performance, and at the same time, to discover the potential opportunities that technology would offer for more effective TBLT courses. Similar to face-to-face TBLT, many factors and conditions influence the quality and quantity of task performance in technology-mediated TBLT; for example, task complexity (Robinson, 2001), corrective feedback (Loewen & Erlam, 2006), task type, teacher factors, learner factors (Oxford, 2006), and so forth. Ellis (2010) points out that for implementation of technology in TBLT environment, there must be a full understanding of the conditions in order to prepare the best design to foster learners’ learning. Accordingly, a well connection is needed among theory, research and practice to set the best condition for the favorable synergy between technology and task-based approaches.

As was mentioned above, one of the factors which influence the task performance is learner factor (Oxford, 2006). It includes, as Oxford (2006, p. 17) states, “different task roles for learners as well as individual learning styles.” Considering the distinction between task, as the work plan, and activity, as the communication which results from the performance of the task, learners interpret the work plan in terms of their own needs, characteristics and motives. So, a single task may result in various activities when performed by different learners or even by the same learners on different occasions and in different contexts (Ellis, 2010). On the other hand, while tasks are done in a technology-mediated environment, more options are available for language learners and this phenomenon could lead to more variant interpretations of a single task as a workplan. This study aimed to investigate the relationship between learners’ individual learning styles and their task performance in technology-mediated context based on lexical density and diversity (types-token ratio), and syntactic complexity and accuracy to find out which learning styles contribute to better task performance.
Literature review

During the last fifty years, it has been witnessed that the application of technology in language learning has been more apparent than ever before and it can be traced back in the development of computer, the Internet and software. In more recent years, a synergy between Task-based Language Teaching and technology is observed. Chapelle believes that the conditions in which language learning take place are changing constantly (Chapelle, 2003). It is undeniable that nowadays language learners have access to vast amount of language materials through the Internet and computers and many of them are using these materials, even when they are not put and predicted in their syllabus. Prensky (2003) believes that today’s world would seem meaningless to students without access to the Internet, computers and digital media.

In a reciprocal approach, Chapelle (2003) points out that TBLT can guide instructors in selection of technology-enhanced language learning sources and materials. This view suggests that technology and TBLT can move together to reach an optimal point where language learners be able to achieve the optimum performance in language learning process. Doughty and Long (2003) also emphasize the reciprocal relationship and interdependence between technology and TBLT. At the same time, they state, TBLT provides a foundation and outline to select proper technological tools and facilities. They, then, point out that the selection among technological options for language learning and teaching purpose must be theoretically and empirically inspired rather than being market-driven.

The aim of all teachers is to make the best environment for their students so that they could learn a new language in the best possible way. To reach this goal, many factors are involved and teachers must be aware of them. These factors correspondingly do exist in technology-mediated language teaching contexts. It doesn’t seem reasonable to conclude that the sole use of technology would bring positive attitudes and better language learning for all the learners with different characteristics, even in a single classroom.

Oxford (2006, p. 9) proposes five dimensions for the task framework: “task type (for example, information-gap), importance of task (low- or high-stake requirement), task complexity (at linguistic or cognitive levels), and teacher factor (task roles for teachers as well as the support teachers give to learners), and learner factor (different task roles for learners as well as individual learning styles).” Among these dimensions, learner factor is the focus of this study. Teachers must know about their students’ characteristics and the best way they can learn a new subject. A single subject might be interpreted and grasped in different ways by different students in a certain language classroom and result in complete chaos.

In the literature, various roles have been assigned to learners such as group participant, monitor, risk taker/innovator, strategy-user, goal-setter, self-evaluator, task-analyzer, and more (for example, Honeyfield, 1993; Oxford, 1990; Richards & Rodgers, 2001). The other learner factor is learner’s individual learning style.

Learning styles are “the attitudes and behaviors which determine an individual’s preferred way of learning” (Honey et al., 1992, p. 1). Keefe (1982, p. 4) defines learning styles as “cognitive, affective, and psychological traits that are relatively stable indicators of how learners perceive, interact with, and respond to the learning environment”. Fenrich (2006) states that instructional designers must consider learners’ learning styles when they are designing certain syllabus and materials in order to achieve the maximum learning state of the students. Research in this area has shown that a match between learning style and instructional materials can lead to learners’ success in learning and performing tasks in order to achieve a particular outcome (Ayersman, 1993; Ghaoui & Janvier, 2004; Price, 2004; Yung-Bin, 1992).

Robin states that “in the immediate future – the next five to ten years – the frontier in language learning and technology will not be found in what program does what better, but rather which students use off-the-shelf technology to best facilitate their own learning in their own learning style” (2007, p. 109). Empirical research suggests that it may be the best to adapt instructional delivery and content to accommodate differences in the ways students learn (Hansen & Stansfield, 1981; Hansen, 1980).

Raschio (1990) examined the role of cognitive style in improving Computer-assisted Language Learning (CALL). In his study he concludes that “for CALL to be effective, we must not only give our students access to computers for reasonable amounts of time; we must also understand their learning strategies and provide exercises that are conducive to their particular cognitive [learning] style” (Raschio, 1990, p. 540).
Canavan (2004) investigated how personalized courses can be delivered to the learner in an adaptive environment. More specifically, he examined how learning style information can be integrated into an Adaptive Hypermedia System to offer increased personalization which will result in better learning. In another study, Shaw (2012) investigated the relationship between learning styles, participation types, and performance in programming language learning supported by online forums. He uses Kolb’s (1999) learning style inventory in his study. He concluded that different learning styles were associated with significantly different learning scores.

**Purpose of the study and research questions**

The aim of this study was to investigate the language produced by learners in a synchronous environment in terms of lexical density and diversity, and syntactic complexity and accuracy in relation to learners’ individual learning styles.

To fulfill the objectives of this study the following questions were raised to be answered:

- Is there any relationship between learning style of the learners and their performance in tasks carried out in the Synchronous Computer-mediated Communication (SCMC) in terms of lexical density and diversity, and syntactic complexity and accuracy?
- Which learning style(s) contribute to better task performance in SCMC?

And the following null hypotheses were formulated for answering the research questions:

H$_{01}$: There is no significant relationship between learners’ learning style and their performance in tasks carried out in SCMC in terms of lexical density and diversity, and syntactic complexity and accuracy.

H$_{02}$: No specific learning style contributes to better task performance in SCMC.

**Material and methods**

**Participants**

In this study participants, initially, consisted of 60 advanced-level Iranian EFL learners. Levy and Stockwell (2006) suggest that learners need to achieve a certain level of language proficiency to deal with synchronous communication; and that is why the advanced-level learners were selected for this study. After employing the learning style inventory, 40 of them with different learning styles were selected and categorized into eight groups with different learning styles defined in Felder-Silverman Learning Style Model. This selection was done in order to have almost equal number of participants for each learning style dimension. Another point was that, the participants with higher tendency toward any learning style dimension were selected.

The participants’ age range was between 20 to 32 and they had at least three years of English learning experience at the time of task performance. They consisted of both females ($N = 23$) and males ($N = 17$) and their education level varied from Bachelor’s to PhD degree. They were selected from EFL learners enrolled at a language institute in Tehran, Iran, which is affiliated to University of Tehran, Faculty of Foreign Languages and Literatures. The participants’ level of proficiency was assumed to be the same as they were in the same level studying Summit 1a (Saslow & Ascher, 2006) and they were classmates for more than four semesters. To be more accurate and definite about participants’ level of proficiency, the researcher had interviews with all of them in separate sessions. This examination had a qualitative nature and no scores were given to participants' performance.

**Instruments**

The instruments used in this study consisted of one learning style questionnaire and one online chat tool. Felder-Soloman’s (1991) Index of Learning Style (ILS) was chosen to be used among different learning style models. The online chat tool was located in a website (www.mohsenhedayati.ir) designed and developed by the researcher for the purpose of the present study.
**Felder-Soloman Index of Learning Style**

There are different learning style inventories in the literature: Kolb (1984), Honey and Mumford (1992) and others. In this study, Felder-Soloman (1991) Index of Learning Style (ILS) was used, which is based on Felder-Silverman (1988) Learning Style Model (FSLSM). The reason for choosing ILS for this study was that “most other learning style models classify learners into a few groups, whereas Felder and Soloman describe the learning style of a learner in more detail, distinguishing between preferences on four dimensions. Another main difference is that ILS is based on tendencies; indicating that learners with a high preference for certain behavior can also act sometimes differently” (Graf et al., 2007, p.3). Carver et al. (1999, p. 34) believe “the Felder Model is most appropriate for hypermedia courseware.” This comment suggests the appropriateness of the ILS for studies in which technology is integrated using different tools in multimedia.

Kuljis and Liu (2005) did a comparison among learning style models and introduced Felder Model as the most appropriate one. In this inventory four dimensions are identified: Active versus reflective learners, sensing versus intuitive learners, visual versus verbal learners, and sequential versus global learners. The inventory consists of 44 items and each learner has a personal preference for each dimension. These preferences are expressed with values between +11 to -11 per dimension, with steps +/-2 (Graf et al., 2007).

Concerning the reliability and the validity of this inventory, a number of studies have reported high reliability and validity for it. Seery et al. (2003), for example, examined the test-retest reliability of this test and reported a high correlation in all the four dimensions (0.804 for Active/reflective dimension, 0.787 for sensing/intuitive dimension, 0.870 for visual/verbal dimension, and .725 for global/sequential dimension; all coefficients are significant at the 0.05 level or better). The internal consistency reliability of ILS which investigates the homogeneity of the items within a test is measured in Zywno’s (2003) study. He uses Cronbach’s coefficient alpha metric and reports almost high homogeneity among the items in all the four dimensions (.60, .70, .63, and .53; all greater than the criterion value of 0.5). In terms of construct validity, which signifies how successfully a certain instrument really measures the theoretical construct, Felder and Spurlin (2005) conducted a study and based on the results confirmed ILS’s construct validity.

**Online chat tool**

In the present study, participants were asked to take part in online chat sessions. Through these chat sessions they were involved in an opinion-gap task performed in the style of free discussion. This online chat accessibility was supplied in a website designed and developed by the researcher for the purpose of the present study. In this web site (www.mohsenhedayati.ir), there is a section for synchronous chat which is designed highly sensitive to the goals of the study.

**Method and procedure**

The procedure of this study consisted of three phases: implementing Learning Style Questionnaire, collecting data from synchronous chat sessions, and data analysis. This study benefited from a descriptive design with a deductive objective.

Since the main part of this study was conducted in a technology-mediated environment using the Internet, the implementation of Learning Style Questionnaire was also done online. When the participants were chosen for the first time, their email addresses were asked and recorded, and afterwards, all the procedure was followed by using the Internet. The ILS questionnaire was emailed to participants in pdf format and they were asked to answer and email it back within one week. Based on the results, the participants were categorized into 8 groups in four dimensions: Active vs. Reflective, Sensing vs. Intuitive, Visual vs. Verbal, and Sequential vs. Global learners. The next step was to hold the synchronous chat sessions for performing the opinion-gap task.

As the main part of the study, the participants were asked to take part in synchronous chat sessions to perform free discussion task. The discussion topics were chosen from the topics in the students’ course book (Summit, 1A), for example: new perspectives into life, musical moods, money matters and etc. This link between the discussion topics
and students course book let them have access to related vocabularies and grammar. There were six online free
discussion sessions, each one lasting for 50 to 70 minutes. These sessions were held twice a week on Mondays and
Thursdays for three consecutive weeks. The researcher took part in the discussions as the facilitator and tried to
prepare the equal opportunities for all the participants to get involved in the discussion. This control let all the
participants share their ideas with others. The researcher introduced the topic of discussion at the beginning of every
chat session and then asked the participants to share their ideas one by one. When one of the students completed
his/her writing, others were invited to comment on his/her ideas.

After the six discussions were conducted, all the texts were recorded and every participant’s produced language was
saved individually to his/her own profile on Microsoft Word software. Every utterance produced by the students in
the chat session was counted as data for analysis. In other words, no random selection was used for data selection.
After six sessions, every student’s profile included around 446-482 words, and 51-62 sentences. The next step began
by analyzing these data based on Lexical density and diversity, and syntactic complexity and accuracy.

The main purpose of this study was to analyze the language produced by language learners in synchronous
computer-mediated communication and examine whether there exist any relationship between their produced
language and their individual learning style. In this regard, students’ produced language was analyzed in terms of
lexical density and diversity at vocabulary level, and syntactic complexity and accuracy at grammar level. For
comparing these measures among groups in the four aforementioned dimensions the descriptive measures of mean
(M) and standard deviation (SD) were used. Then, independent samples t-test was conducted to check if the
differences between the groups were significant.

Lexical density was measured by calculating the proportion of lexical (content) items to the total number of words in
the text (Ure, 1971). Lexical items involved nouns, verbs, adjectives, and most of the adverbs. In terms of lexical
diversity, texts were analyzed to find out how many different words are used by a single student in their produced
language. To calculate this measure, the WordSmith Tools (Scott, 2012) software was employed.

At grammar level, two measures of syntactic complexity and error-free C-units were used. Regarding syntactic
complexity, the Average Sentence Length (ASL) index was utilized. To determine the average sentence length, the
average number of words used in every sentence was calculated. This measurement was also done using WordSmith
Tools (Scott, 2012) software. For the second measure, the percentage of error-free C-units in learners’ produced
language was calculated. The rationale behind choosing C-units is that they include partial sentences as well as
complete sentences. In contrast, T-units only include sentences that can stand alone. Since many utterances in online
communication include partial sentences, using c-units, all the utterances could be analyzed. It’s worth mentioning
that, 50 percent of the analyses (i.e., lexical density and error-free c-units) which were done by the researcher were
reanalyzed by an experienced Ph.D. student in TEFL.

Results

Learners’ learning style frequency

In the first phase of this study, the researcher identified language learners’ individual learning style using ILS and
categorized them into eight groups in four dimensions (see Tables 1, 2, 3, and 4).

| Table 1. Frequency of participants, first dimension (active/reflective) |
|-------------------------|----------------|----------------|----------------|----------------|
|                         | Frequency | Percentage | Valid percentage | Cumulative percentage |
| Valid                   | Active     | 21          | 52.5          | 52.5           | 52.5             |
|                         | Reflective | 19          | 47.5          | 47.5           | 100              |
| Total                   | 40         | 100         | 100           |                |                  |

| Table 2. Frequency of participants, second dimension (sensing/intuitive) |
|-------------------------|----------------|----------------|----------------|----------------|
|                         | Frequency | Percentage | Valid percentage | Cumulative percentage |
| Valid                   | Sensing   | 18          | 45             | 45             |
|                         | Intuitive | 22          | 55             | 55             |
| Total                   | 40        | 100         | 100            |                |

348
Table 3. Frequency of participants, third dimension (verbal/visual)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percentage</th>
<th>Valid percentage</th>
<th>Cumulative percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>13</td>
<td>32.5</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Visual</td>
<td>27</td>
<td>67.5</td>
<td>65</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Frequency of participants, fourth dimension (sequential/global)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percentage</th>
<th>Valid percentage</th>
<th>Cumulative percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>17</td>
<td>42.5</td>
<td>37.5</td>
<td>37.5</td>
</tr>
<tr>
<td>Global</td>
<td>23</td>
<td>57.5</td>
<td>62.5</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Lexical density and diversity analysis

At vocabulary level, the participants’ produced language was analyzed using two measures of Lexical Density and Lexical Diversity (see Table 5). Lexical Density was calculated by the formula developed by Ure (1971):

\[
LD = \frac{\text{Lexical items}}{\text{Total number of words}} \times 100
\]

As the purpose of this study, the mean lexical density of the language produced by participants in every group was compared with the other group’s mean in every one of the dimensions using independent samples t-test and the following results were gained. The same process was performed for lexical diversity, but this measure was calculated by Wordsmith software (2012).

In the first dimension, it was revealed that the reflective learners outperformed the active ones in terms of lexical density (by the overall mean of 45.83 in comparison to the overall mean of 41.69). An independent samples t-test was conducted to compare the lexical density scores for active and reflective learners. As shown in Tables 5 and 6, there was a significant difference in the mean lexical density of active learners \((M = 41.69, SD = 4.39, N = 21)\) and reflective learners, \(M = 45.83, SD = 5.73, N = 19; t(38) = 1.71, P = .050\) (two-tailed). These results showed that reflective learners produced lexically denser sentences in comparison to active learners.

Table 5. Descriptive statistics for students’ performance at vocabulary level

<table>
<thead>
<tr>
<th>Dimension 1</th>
<th>Lexical density</th>
<th></th>
<th></th>
<th>Lexical diversity</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>Sig.</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Active</td>
<td>21</td>
<td>41.6962</td>
<td>4.39676</td>
<td>.050</td>
<td>21</td>
<td>56.6576</td>
</tr>
<tr>
<td>Reflective</td>
<td>19</td>
<td>45.8395</td>
<td>5.73873</td>
<td>.050</td>
<td>19</td>
<td>60.5979</td>
</tr>
<tr>
<td>Dimension 2</td>
<td>Sensing intuitive</td>
<td>18</td>
<td>40.0522</td>
<td>3.53534</td>
<td>.011</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22</td>
<td>44.0286</td>
<td>5.44868</td>
<td>.011</td>
<td>22</td>
</tr>
<tr>
<td>Dimension 3</td>
<td>Verbal</td>
<td>13</td>
<td>38.3862</td>
<td>2.57768</td>
<td>.000</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Visual</td>
<td>27</td>
<td>44.0944</td>
<td>4.90996</td>
<td>.000</td>
<td>27</td>
</tr>
<tr>
<td>Dimension 4</td>
<td>Sequential</td>
<td>17</td>
<td>43.9918</td>
<td>5.14807</td>
<td>.050</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>23</td>
<td>40.9439</td>
<td>4.65938</td>
<td>.050</td>
<td>23</td>
</tr>
</tbody>
</table>

The mean lexical density was also different in the other three dimensions, and these differences were all significant (see, Tables 5 and Figure 1). In the second dimension, intuitive learners \((M = 44.02, SD = 5.44, N = 22)\) outperformed the sensing learners \((M = 40.05, SD = 3.53, N = 18)\) and this difference was significant; \(t(38) = 2.66, P = .011\) (two-tailed). In the third dimension, the results also indicated that visual learners \((M = 44.09, SD = 4.90, N = 27)\) produced lexically denser sentences in comparison to verbal learners \((M = 38.38, SD = 2.57, N = 13)\). In the last dimension, sequential learners \((M = 43.99, SD = 5.14, N = 17)\) outperformed the global learners \((M = 40.94, SD = 4.65, N = 23)\). The significance levels of these differences were calculated using independent-samples t-test and results are reported in Table 6.
The results for the lexical diversity analysis also indicated significant difference between groups in the four dimensions (see Table 7 and Figure 2). The independent samples t-test was also conducted at this level. The results showed that reflective learners ($M = 60.59$, $SD = 7.31$, $N = 19$) used more variant words in their sentences, in comparison to active learners ($M = 56.65$, $SD = 8.87$, $N = 21$); $t(38) = 1.36$, $P = .050$ (two-tailed). In the second dimension, the results indicated that intuitive learners ($M = 59.52$, $SD = 8.51$, $N = 22$) used greater variety of words in their produced language comparing to sensing learners ($M = 54.14$, $SD = 6.57$, $N = 18$); $t(38) = 2.19$, $P = .034$ (two-tailed). In the third dimensions, visual learners ($M = 59.25$, $SD = 6.94$, $N = 27$) had higher lexical diversity rate than verbal learners ($M = 52.64$, $SD = 8.70$, $N = 13$); $t(38) = 2.59$, $P = .013$ (two-tailed). In the fourth dimension, Sequential learners ($M = 59.17$, $SD = 8.66$, $N = 17$) used more different lexical words in the language they produced comparing to global learners ($M = 56.31$, $SD = 7.71$, $N = 23$); $t(38) = 1.71$, $P = .050$ (two-tailed).
Table 7. Independent samples t-test for Lexical diversity means

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Active</th>
<th>Reflective</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean difference</th>
<th>Std. error difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension 1</td>
<td>-1.363</td>
<td>38</td>
<td>.050</td>
<td>4.94028</td>
<td>2.58721</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension 2</td>
<td>Sensing</td>
<td>Intuitive</td>
<td>-2.199</td>
<td>38</td>
<td>.034</td>
<td>-5.38753</td>
<td>2.45030</td>
</tr>
<tr>
<td>Dimension 3</td>
<td>Verbal</td>
<td>Visual</td>
<td>-2.592</td>
<td>38</td>
<td>.013</td>
<td>-6.60345</td>
<td>2.54714</td>
</tr>
<tr>
<td>Dimension 4</td>
<td>Sequential</td>
<td>Global</td>
<td>1.715</td>
<td>38</td>
<td>.050</td>
<td>2.85962</td>
<td>2.60064</td>
</tr>
</tbody>
</table>

Syntactic complexity and accuracy analysis

Data was first analyzed at vocabulary level, in the first phase of the study, and then the same data was investigated at grammar level. To analyze the learners’ produced language at grammar level, two measures of Average Sentence Length (ASL) and percentage of error-free c-units (EFU) were used (see Table 8). The ASL was calculated using Wordsmith tool (2012), and the calculation of mean percentage of error-free c-units was accomplished by the researcher. The results showed that, in general, learners in some of the groups of learning style model performed differently at grammar level in comparison to vocabulary level.

At grammar level reflective learners showed better performance in terms of both mean average sentence length (see Table 9 and Figure 3) and mean percentage of error-free c-units (see Table 10 and Figure 4). The results of independent samples t-test showed that reflective learners (M = 9.12, SD = 2.89, N = 19) produced longer sentences in comparison to active learners (M = 7.40, SD = 2.38, N = 21); t(38) = 2.06, P = .046 (two-tailed). In the second dimension, the results of independent samples t-test indicated that sensing learners (M = 9.50, SD = 3.09, N = 18) produced longer sentences in comparison to intuitive learners (M = 7.98, SD = 2.47, N = 22); and this difference between means was significant; t(38) = 1.58, P = .050. These results showed that sensing and intuitive learners had different performance at vocabulary and grammar level. At vocabulary level, intuitive learners outperformed the sensing learners, but at grammar level, the opposite was true.

In the third dimension, visual learners outperformed verbal learners; the same was witnessed at vocabulary level. The results of the independent samples t-test revealed a significant difference between the mean of average sentence length of visual learners (M = 10.22, SD = 2.64, N = 27) and the verbal learners (M = 8.20, SD = 3.07, N = 13); t(38) = 1.02, P = .040 (two-tailed). In the last dimension, results were not the same as at vocabulary level; global learners had better performance than sequential learners at grammar level. Global learners (M = 9.10, SD = 2.67, N = 23) produced longer sentences than sequential learners (M = 7.01, SD = 2.42, N 17), and this difference between means was shown to be significant; t(38) = 2.53, P = .015.

Table 8. Descriptive statistics for students’ performance at grammar level

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Active</th>
<th>Reflective</th>
<th>Mean</th>
<th>SD</th>
<th>Sig.</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension 1</td>
<td>21</td>
<td>19</td>
<td>7.4014</td>
<td>2.38714</td>
<td>.046</td>
<td>21</td>
<td>43.952</td>
<td>6.0578</td>
<td>.030</td>
</tr>
<tr>
<td>Dimension 2</td>
<td>18</td>
<td>22</td>
<td>9.5033</td>
<td>3.09680</td>
<td>.050</td>
<td>18</td>
<td>46.211</td>
<td>6.9377</td>
<td>.040</td>
</tr>
<tr>
<td>Dimension 3</td>
<td>13</td>
<td>27</td>
<td>8.2038</td>
<td>3.07325</td>
<td>.040</td>
<td>13</td>
<td>44.923</td>
<td>6.8975</td>
<td>.030</td>
</tr>
<tr>
<td>Dimension 4</td>
<td>17</td>
<td>23</td>
<td>7.0188</td>
<td>2.42346</td>
<td>.015</td>
<td>17</td>
<td>42.876</td>
<td>4.9150</td>
<td>.024</td>
</tr>
</tbody>
</table>
At accuracy level, the data was analyzed by percentage of error-free c-units measure to investigate the accuracy level of the produced language by the learners. Independent samples $t$-test was used to compare the means between the groups in the four dimensions. The results of the analysis revealed a significant difference between the means of the error-free c-units' percentage between the groups. In the first dimension, the results showed that reflective learners ($M = 47.95$, $SD = 5.96$, $N = 19$) were more accurate than active learner ($M = 43.95$, $SD = 6.05$, $N = 21$) and this difference was significant; $t(38) = 1.57$, $P = .030$. In the second dimension, sensing learners ($M = 46.21$, $SD = 6.93$, $N = 18$) were more accurate than intuitive learners ($M = 44.69$, $SD = 5.44$, $N = 22$); $t(38) = 1.77$, $P = .040$. In the third dimension, visual learners ($M = 46.59$, $SD = 5.84$, $N = 27$) outperformed the verbal learners ($M = 44.92$, $SD = 6.89$, $N = 13$) in terms of accuracy; $t(38) = 1.32$, $P = .030$. In the fourth dimension, global learners ($M = 47.22$, $SD = 6.36$, $N = 23$) were more accurate than sequential learners ($M = 42.87$, $SD = 4.91$, $N = 17$); $t(38) = 2.34$, $P = .024$ (two-tailed).
Table 10. Independent samples t-test for error-free

t-test for equality of means (Error-free c-units)

<table>
<thead>
<tr>
<th>Dimension 1</th>
<th>Active</th>
<th>Reflective</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.576</td>
<td>38</td>
<td>.030</td>
<td>-4.0003</td>
<td>1.9041</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimension 2</th>
<th>Sensing</th>
<th>Intuitive</th>
<th>1.774</th>
<th>38</th>
<th>.040</th>
<th>1.5157</th>
<th>1.9574</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension 3</td>
<td>Visual</td>
<td>Verbal</td>
<td>-1.322</td>
<td>38</td>
<td>.030</td>
<td>-2.6732</td>
<td>2.0925</td>
</tr>
<tr>
<td>Dimension 4</td>
<td>Sequential</td>
<td>Global</td>
<td>-2.344</td>
<td>38</td>
<td>.024</td>
<td>-4.3496</td>
<td>1.8557</td>
</tr>
</tbody>
</table>

**Discussion**

Felder and Soloman (1991) described the characteristics of the learners with any of the learning styles in FSLSM. They suggested that, in the first dimension, active learners learn best by doing something physically and prefer group working, while reflective learners prefer to think about something comprehensively and they tend to work alone. The results of this study showed that reflective learners produced lexically denser sentences in comparison to active learners; this was also witnessed in terms of lexical diversity. This difference was more obvious at grammar level where reflective learners produced longer sentences with higher percentage of error-free c-units. The longer sentences with higher percentage of error-free c-units can denote a relationship between reflective learners’ thoughtfulness and their better performance at grammar level in SCMC. In other words, it can be discussed that reflective learners sustained their thoughtfulness even in the fast nature of synchronous online chat rooms and tried to be more accurate rather than proficient. Reflective learners showed that they are not affected negatively with the time-pressure existed in the online chat. The other point is that, online discussions are assumed to be group works done by all the students and in this regard, students who prefer group work (Active learners) should have better performance. But the results of this study showed that reflective learners, who prefer working alone, had better performance.

Regarding the second dimension, Felder and Soloman (1991) stated that sensors have tendency toward being more practical and careful in comparison to intuitors who deemed to be more innovative. In this study, sensors produced longer sentences with higher percentage of error-free c-units, while intuitors outperformed sensors at vocabulary level by producing lexically denser and more variant sentences. These results suggest that, learners with sensing learning style attempted to follow well-established rules of syntax to produce accurate utterances. Sensors also used fewer abbreviations in comparison to intuitors. Intuitive learners benefited from their innovation ability and produced more abbreviated utterances among which some were coined by themselves at the moment of conversation. The following example shows the use of abbreviations coined by an intuitive learner (S1):

*S1: Even most lenient prnts are extremely strict in some mtrrs.*

This example shows that intuitive learners are more successful in delivering their intended meaning in the shortest possible way. On the other hand, they might neglect spelling or grammatical rules and produce erroneous sentences.

Visual learners had lexically better performance than verbals and this was also witnessed at grammar level. Since visual learners prefer learning by pictures and analogous stuffs (Felder & Soloman, 1991), they benefited significantly from emoticons in the chat sessions. This opportunity let them convey their emotions and feelings using the small emoticons rather than verbal options. In some cases the number of emoticons denotes the degree of their specific feeling; as in the following example produced by a visual leaner (S1):

*S1: we can observe indirectly & slap directly :-D ·:-D ·:-D*

The online chat environment was not highly welcomed by verbal learners who prefer to be engaged in defined and rule-governed discussions. They mostly commented on the sentences written by the teacher rather than other students.
In the last dimension sequential learners had better performance at vocabulary level comparing to global learners. In contrast, at grammar level, global learners outperformed sequential learners. Felder and Soloman (1991) claimed that sequential learners follow linear steps for gaining new information and are able to make logical relationship among pieces of information. In other words, they have an analytic approach toward finding solutions for the existing problems, however, global learners, as the term “global” denotes, have synthetic approach toward solving problems and they need to get a whole picture of the new information to be able to deal with. These features, in my opinion, may not justify the existence of the relationship between sequential learners’ outperformance at vocabulary level, and respectively global learners’ outperformance at grammar level.

Based on the findings of this study, the first null hypothesis which assumed that “there is no significant relationship between learners’ learning style and their performance in tasks carried out in SCMC in terms of lexical density and diversity, and syntactic complexity and accuracy,” is rejected. The second research question asked whether any learning style(s) contribute to better task performance in synchronous computer-mediated communication (SCMC). The findings of the present study revealed that two groups of learning styles, reflective learners and visual learners, outperformed other groups at both vocabulary and grammar levels. In other words, reflective learners, as well as visual learners produced lexically denser sentences with higher variation of words. At the same time, they produced longer sentences with higher percentage of error-free c-units. Based on these findings the second null hypothesis was also rejected, where reflective learners as well as visual learners outperformed others lexically and syntactically. This difference in performance suggests that students with varying learning styles go through various learning experiences while using computer tools for communication.

Conclusions

Language learners’ individual learning styles determine the pattern through which they deal with language input, as well as output. This research showed that when language learners find consistency between their preferred way of learning and the form of presented materials, they show higher access to their language abilities and display improved performance. In other words, individual differences can be the sources of performance variations in a language class. The results suggest that we should not expect the same performance from all the students in a technology-mediated language learning course until we have comprehensive information about their preferred way of learning by knowing their individual learning styles which determines looked-for ways of learning. This information will let both students and teachers to be more accurate in addressing the right area of discrepancy and divergence. This study confirms the idea that mere use of technology in language education is not a panacea which would results in increased language learning for all students.

Based on these findings, it is worth studying the feasibility of designing a technology-enhanced syllabus which corresponds to language learners’ individual learning style, while sustaining the ultimate goals of the course.

At the time of the performance of this research there were some limitations which did not let the researcher include more variables to investigate the issue more comprehensively. For instance, language learners’ setting at the time of online chat were not specified or homogenized. On the other hand, the setting heterogeneity of participants in an online chat room is inevitable in real-life situation. The other limitation was the varying number of participants in each group. While the number of participants could not be increased due to group management parameters, some learning styles are observed less than others in real-life situation and it was not possible to find equal number of participants for every group at the moment of study. In addition, task type was not included and investigated in this study, which could probably affect the results. The only task type used in this study was opinion-gap task; while other tasks like jigsaw and information-gap can be included. This study also can be performed in an asynchronous mode to see if the same results will be achieved; this study was conducted in a synchronous mode.

References


Yung-Bin, B. L. (1992). Effects of learning style in a hypermedia instructional system. In *Proceedings of selected research and development presentations at the convention of the Association for Educational Communications and Technology* (pp. 506-508).

Has Research on Collaborative Learning Technologies Addressed Massiveness?
A Literature Review

Kalpani Manathunga1* and Davinia Hernández-Leo2

1ICT Department, Universitat Pompeu Fabra, Barcelona, Spain // 2Serra Húnter Fellow //
kalpani.manathunga@upf.edu // davinia.hernandez@upf.edu

*Corresponding author

(Submitted August 4, 2014; Revised March 16, 2015; Accepted March 25, 2015)

ABSTRACT

There is a growing interest in understanding to what extent innovative educational technologies can be used to support massive courses. Collaboration is one of the main desired elements in massive learning actions involving large communities of participants. Accumulated research in collaborative learning technologies has proposed and evaluated multiple models and implementation tools that engage learners in knowledge-intensive social interactions fostering fruitful learning. However, it is unclear to what extent these technologies have been designed to support large-scale learning scenarios involving arguably massive participation. This paper contributes with a literature review that aims at providing an answer to this question as well as offering insights about the context of use, characteristics of the technologies, and the types of activities and collaboration mechanisms supported. The main results point out that till 2013 the level of massiveness considered in top scientific journal papers on collaborative learning technologies was low, the scenarios studied were predominantly contextualized in co-located higher education settings using Learning Management Systems, the most common activities considered were open and structured discussion, followed by peer assessment and collaborative writing, and the most broadly used mechanism to foster fruitful collaboration was group formation following diverse policies.

Keywords
Literature review, Educational technologies, Collaborative learning, Large classes, Massive courses

Introduction

The interest in educational technologies for massive numbers of learners has recently increased because of the impact that Massive Open Online Courses (MOOCs) are having in the Media and the Society (Sonwalkar, Wilson, Ng, & Sloep, 2013). This impact is shaping a turning point in educational technologies research as it offers an excellent opportunity for the adoption of previous research achievements while creating new scalability research challenges for massive teaching, learning and assessment models (Kay, Reimann, Diebold, & Kummerfeld, 2013). Aligned with existing research evidences, MOOC initiatives recognize the importance of social interaction among learners. Many of them incorporate activities based on discussions and peer-review assessments (Tsai & Wong, 2013). The potential for different pedagogies and collaborative learning methodologies in massive class teaching is prospective and highly concerned. Yet, we envision that this potential is still in its infancy though undoubtedly relevant open discussions and peer-assessments are the main two examples of collaborative learning techniques with current practice in massive courses (Kay et al., 2013; Sonwalkar et al., 2013; Tsai & Wong 2013).

Collaborative learning techniques support the construction of joint knowledge and sharing of meanings by means of fostering potentially effective social interactions (Dillenbourg, 1999). Accumulated research in Computer-Supported Collaborative Learning (CSCL) has proposed and evaluated multiple models and implementation tools that engage learners in knowledge-intensive social interactions (debate, conflict resolution, artifact co-design, mutual explanation, etc.) with identified significant learning outcomes (Dillenbourg, 1999; Stahl, Koschmann & Suthers, 2006). These models involve the application of collaboration-triggering mechanisms such as group formation according to specific policies, role allocation and rotation, distribution of knowledge, etc. and the use of diverse collaboration spaces (shared boards, wikis, etc.) and implementing communication and coordination mechanisms (flow control, group awareness, etc.). However, the standout body of CSCL research is mostly known for its contributions focused on supporting small groups of learners (Stahl, Law, & Hesse, 2013). And the research around scalable collaborative learning approaches, technologies and issues for large classrooms or large learning communities is scattered across scientific publications without explicitly embracing a comprehensive visible body of knowledge.
This paper contributes with a systematic literature review (Kitchenham, 2004; Webster & Watson, 2002) synthesizing a framework that explains existing insights and gaps in the context of applying collaborative learning aimed at massive or large groups. This framework will serve as a foundation for advancing knowledge and uncovered areas (Webster & Watson, 2002) where further research in above aspects could be conducted accordingly. Hence, the rationale for the paper is not to identify CSCL as a branch of MOOCs providing collaboration aspects, but to understand to what extent previous research in CSCL has involved in the design or/and use of technologies suitable to support massive or large-scale participation. The ultimate aim is to characterize which technologies and approaches could be potentially used in MOOCs (or, more generally, in massive learning actions) to support collaboration - because its use with relatively large learner communities has been proved and studied. As a secondary aim, the paper also discusses challenges and promising avenues emerging from the literature review.

Therefore, a first aim of the analysis is identifying the context types for research works that have considered arguably (or potentially) massive / large quantities of learners in the concerned technology-supported collaborative learning scenarios. Then, the concrete focus is on understanding the types of tasks or activities supported by collaborative learning technologies in those scenarios, as well as the types of mechanisms and technological facets considered by these technologies to support collaboration. A systematic approach is followed consisting of stages as (1) identify research objectives (2) search articles (3) filtration and evaluation of data set (4) coding and analysis (5) interpretation of results obtained. The coding of the data is done using a qualitative data analysis tool (Atlas.ti), whose features facilitate researchers a systematic management and coding of text instances in articles. As the research objectives, following specific research questions were formulated:

In research involving arguably or potentially massive technology-supported collaborative learning environments,
- RQ1: to what extent the scenarios considered are massive?
- RQ2: what are the types of educational sectors and settings considered?
- RQ3: what types of activities are proposed?
- RQ4: what collaboration mechanisms are implemented?
- RQ5: which are their technological facets?

Educational sectors or levels (from primary to adult education and informal learning) and the types of settings (co-located, remote, in physical or virtual spaces) characterize the context of the learning scenarios in which research on collaborative learning technologies have been framed. The literature review will provide insights about in which contexts these technologies have been applied with many learners. As mentioned above, collaboration environments have been proposed for a number of diverse activities (from debates to product co-development), the review will provide light about to what extent these environments have been used in massive situations. Group formation following specific policies and distribution of roles and knowledge are design techniques used in pedagogical methods and technologies (such as collaboration scripts) seeking potentially fruitful social interactions (Dillenbourg, 2002; Hernández-Leo et al., 2006; Dillenbourg, Järvelä, & Fischer, 2009). A potential research question is whether these approaches have been designed for massive scales also.

A realistic educational scenario could have multiple educational tools and technologies involved, including Learning Management Systems (LMS), generic tools, devoted tools, pervasive and ubiquitous devices (Harrer, Pinkwart, McLaren, & Scheuer, 2008; Suo, Miyata, Morikawa, Ishida, & Shi, 2009; Calvo, O’Rourke, Jones, Yacef, & Reimann, 2011). Therefore, inquiring about the technological platforms considered, the interactions between tools and to what extent they are seamlessly connected also has a scientific interest corresponding to massive learning situations. Also from the technological facets perspective, seamless learning implies certain type of interoperability between tools or an enabling technology that acts as a mediator to allow learners to feasibly switch and flow between diverse physical and virtual spaces (Chan et. al, 2006; Pérez-Sanagustín et al., 2012).

The remainder of the paper is structured as follows. Next section details the methodology followed, including the procedure applied to identify possible similar reviews, the search criteria for the literature considered and the method of analysis. Then, a results section is organized as subsections, based on the structure of the research questions. This is followed by a discussion section, which explains requiring concerns on the research aspects with prevailing challenges. The paper concludes with a conclusion of the main findings.
Methodology

Originality of the literature review

A first phase was devoted to identify if a similar literature review attempt was already available. A search clause was formulated including set of keywords denoting the focus of the targeted research topic (Webster & Watson, 2002). The search clause comprised of “review” or “state of the art” or “bibliography” or “survey” (as the nature of the targeted contribution) covering key aspects like “learning” or “education” in “collaborative” or “cooperative” “computer” or “technologically” supported environment targeting “large” or “massive” classes or even “communities” or “MOOCs.”

The resulting search clause with the complete criterion for title search was: (review OR state of the art OR bibliography OR survey) AND (education* OR learning) AND (collaborat* OR cooperat*) AND (large OR massive OR MOOC OR communit*) AND (comput* OR technolog*).

Databases including IEEE Xplore, Web of Science, ACM, Scopus, SpringerLink, ScienceDirect and GoogleScholar were considered for the search since they cover a significant wide range of Computer Science, Education and interdisciplinary scientific publications. The comprehensive search query returned 0 results indicating that this specific topic had not been studied so far. There were no journal articles (either peer reviewed or not), no conference publications neither any text available for that specific topic at the time of writing this article as search query did not present any specific time period.

Literature selection

The next iteration of the search process was to seek the relevant literature to consider in the review (Webster & Watson, 2002). Based on the above formed research questions, a series of keywords were recognized and search was extended up to title, abstract and keywords. Subset of keywords were (education* OR learning) AND (collaborat* OR cooperat*) AND (large OR massive OR MOOC OR communit*) AND (comput* AND technolog*).

For this literature selection Scopus was selected as the database source, given its wide scope that includes the relevant Educational Technology publications (Falagas, Pitsouni, Malietzis, & Pappas, 2008; Chou, 2012), such as those ranked in the “top peer reviewed journals with high impact factors” by Google Scholar (under “Engineering and Computer Science” or “Social Sciences” sub-field “Educational Technology”). Also Scopus gives the facility of maintaining lists of selected papers and provides a graphical view of publications over time. To select the most appropriate and accurate work, few limitations/criteria were implied; such as sources being either peer reviewed journal articles or conference proceedings, published date fallen between 2000 up to December, 2013 and concerned fields are being physical sciences and social sciences while eliminating life and health sciences as they are not related directly with technology enhanced learning. We decided to include aforementioned timespan (starting in 2010) since research on technology supporting collaborative learning was emerging at that time (Dillenbourg, 1999; Stahl, Koschmann, & Suthers, 2006).

Initial screening in Scopus resulted of 6514 papers containing above key terms in topics, abstracts or keywords in only peer-reviewed journal articles and conference proceedings. 3118 articles of them were journal papers. The temporal distribution of the publications is shown in Figure 1. Years from 2007 to 2011 experienced the highest rates of increase in terms of numbers of publications in the topics, reaching certain equilibrium as of 2011.

Out of the articles potentially relevant it was required a solid logical filtration for the final selection. Two researchers participated in the final selection phase. As for the first stage, top ranked educational technology journals according to an intersection of the rankings in ISI Journal Citation Report (ISI, 2013) and Google Scholar were concerned and this criterion narrowed down the count to 243 journal articles. During the next stage of article filtration, 100 articles (out of the 243) were chosen as relevant and appropriate by considering the topic, abstract and keywords of each article. The topics of those not considered relevant were diverse: e.g., misplaced topic, not addressing collaboration among learners but among teachers or other stakeholders, unclear role of technology supporting learning activities, etc. A summary of the selected papers (by journal) is presented in Table 1.
Selected articles were coded using Atlas.ti, a qualitative data analysis tool (see http://www.atlasti.com/index.html) that helped the categorization process and also the exploration and annotation of articles while providing convenient navigation among the article collection. Most articles were experimental studies from which the two researchers could capture code instances by reading through the methodologies, experimental details and their pedagogical approaches. In this qualitative analysis process, thematic categories or code structure was formed considering the research questions and related classifications widely recognized in the educational and the CSCL fields. Besides, since qualitative data analysis involves the identification and interpretation of themes in textual data, additional codes or sub categories emerged during the analysis process (Kitchenham, 2004; Webster & Watson, 2002). Table 2 collects the structure (or tree) of codes used. The root categories relate directly to the research questions and level one to its main characteristics (see for instance the codes for RQ2 or RQ4). The categories in level two add another layer to the analysis. Most subcategories in this level were formulated at the beginning of the research process, in alignment with the research questions and considering existing categorizations. Examples are types of educational sectors or levels (RQ2), collaboration mechanisms (or CL mechanisms) - either designed by practitioners or socially by students - (RQ4) or “CL activity” type being open or structured discussions or peer assessment (RQ3) or the technological facets used (RQ5) (Dillenbourg, 2002; Hernández-Leo et al., 2006; Dillenbourg et al., 2009). Another set of codes emerged during the analysis, in particular, additional types of activities found in some articles like
game/role playing or collaborative presentations (RQ3) and the subcategories defined in the quotient of learners derived from the specific sample sizes revealed in articles (RQ1).

<table>
<thead>
<tr>
<th>Level two</th>
<th>Level one</th>
<th>Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 100</td>
<td>Co-located (same setting)</td>
<td>Quotient of learners (RQ1)</td>
</tr>
<tr>
<td>Between 100 and 1000</td>
<td>Remote locations (across settings)</td>
<td>Educational setting profile (RQ2)</td>
</tr>
<tr>
<td>More than 1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td>CL activity (RQ3)</td>
</tr>
<tr>
<td>Secondary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open Discussions</td>
<td>Designed by teacher</td>
<td>CL mechanism (RQ4)</td>
</tr>
<tr>
<td>Structured Discussions</td>
<td>Decided by learner</td>
<td></td>
</tr>
<tr>
<td>Peer assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game playing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborative writing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborative presenting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group formation</td>
<td>Platforms and tools</td>
<td>Technological facets (RQ5)</td>
</tr>
<tr>
<td>Role distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouping based on previous performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMS or LMS-embedded tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generic tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completely new tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No seams</td>
<td>Addressing seams</td>
<td></td>
</tr>
<tr>
<td>Seamless</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seamful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seamless</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Regarding RQ5, “platforms and tools” is considered as a code and as sub-codes the categories considered were “LMS” or “LMS-embedded tools” or any “general tool” such as social media like Facebook or Blogger, communication tools like Skype or NetMeeting, etc. When researchers had suggested a complete new tool, the article was coded as a “complete new tool.” Seamless support is another aspect concerned in RQ5. Hence the articles were also analyzed highlighting such ideas as “seamless” and “seamful” (i.e., inflexible switch between learning spaces), if those had looked into being pervasive and the remaining as “no seams” where there is not a need of being seamless.

Results

Level of massiveness and educational setting profiles

The level of massiveness (research question RQ1), in terms of number of learners participating in activities supported by collaborative learning environments, found in the scientific journal papers selected was relatively low if compared to the many learners involved in MOOCs (Kay et al., 2013; Sonwalkar et al., 2013). 55% from the concerned sample have experimented with less than 100 of participants. 37% involved more than 100 students, yet these did not exceed 1000 (Table 3). Only 8% papers found with a quantity of learners larger than 1000 and this could be recognized as a practical issue observed with experiments addressing large class learning as mentioned in most of the papers. On the other hand several researches mentioned that having larger sample sizes of learners involved in the studies would be interesting from a quantitative research perspective (Walta & Nicholas, 2013; Junco, Elavsky, Heiberger, 2013; Williams, Lewis, Boyle, & Brown, 2011; Ferriman, 2013).

Regarding the types of educational sectors (RQ2), 66.4% of the researchers had experimented collaborative learning technological models and tools with students pursuing higher studies like undergraduate courses, post-graduate work
or any vocational studies such as training teachers, nurse, etc. 20.7% found addressing attempts implementing computer supported collaborative activities for primary or secondary schools. 7.7% articles were evident as attempts to embrace technology into adult learning context and another. 67.1% of research had been conducted in co-located situations (same location, physical space) rather than across physical and remote/virtual locations. A summary of the educational sectors by type of setting is provided in Figure 2. Majority (> 50%) from the articles that had experimented with more than 100 participants had been implemented upon higher educational set up and very few attempts were observed for primary or secondary education while another few targeting at vocational training.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Number of learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>Less than 100</td>
</tr>
<tr>
<td>37</td>
<td>Less than 1000, but greater than 100</td>
</tr>
<tr>
<td>8</td>
<td>Greater than 1000</td>
</tr>
</tbody>
</table>

Table 3. Quotient of learners

Only 5.2% was observed as informal learning contexts; e.g., where learners share knowledge according to their co-interests in a community (Huang, Yang, Huang, & Hsiao, 2010; Li et al., 2011) These attempts had shown the potential in promoting further collaboration in informal contexts as discussed by Li et al. (2011) by providing a sustainable teamwork platform for researchers to aid in content sharing and managing knowledge resources to build up scholarly communities of knowledge. Further they envision on scaling up to larger learner communities using teamwork platform, SWiCLE to support lifelong learning (Li et al., 2011). Also Huang et al. (2010) article enlightens construction of mobile collaborative learning networks on top of an existing web based platform by the provision of intelligent grouping services to recommend users learning partners of the same interests and specialities. These studies inspire innovate opportunities of using personal data derived from social networks or any other informal contexts for new pedagogical approaches when scaling up collaborative learning strategies, for example in the context of MOOCs.

**Figure 2. Percentages of educational sectors by settings**

Types of collaborative learning activities supported

The third research question RQ3 refers to the type of collaborative activities attempted in the literature aiming fruitful social interactions. The analysis derived six main types of activities considered in “potentially massive” collaborative learning environments as using virtual discussion boards like forums and blogs, peer assessments, games and collaborative writing spaces like Wiki. Figure 3(a) provides a summary as presented in the literature and Figure 3(b) shows the variation of collaborative activities across educational sectors. Discussions could be considered as the most prevalent activity category with highest number of attempts in literature and generally every activity category has higher values for students pursuing higher studies such as undergraduate or postgraduate students.
60.9% papers have implemented some sort of discussion forums; either structured or open discussions among learners in order to maintain collaboration while another 39.1% have attempted different types of techniques to implement a collaborative learning environment. The highest percentage (38.5%) is inspired by open discussions among learners with no interference of the instructor or a specified structure (Shaw, 2013; Oliveira, Tinoca, & Pereira, 2011; Noroozi, Teasley, Biemans, Weinberger, & Mulder, 2013). 22.4% of articles were promoting structured discussion, through scaffolding by practitioners along the discussion activity or via the implementation of collaborative inquiry models that propose the use of epistemic categories or distribution of roles that condition participation in discussions (Walta & Nicholas, 2013; Ferriman, 2013; So, Seah, & Toh-Heng, 2010; Schellens et al., 2007; Gerosa et al., 2010; Ligorio, Talamo, & Pontecorvo, 2005). Ferriman (2013) had let teachers to provide constructive feedback in the middle of an essay writing activity to guide learners and based on teacher feedback, students posted content to the virtual discussion threads ensuring quality content. Walta and Nicholas (2013) use a Community of Inquiry model in which learners participated in structured discussions like designated blogs, journal spaces and small group tutorial discussions along with open discussions via LMS promoting collaboration and reflective learning. Another study uses collaborative inquiry-centred pedagogy (So et al, 2010) providing opening cues like “My theory,” “I need to understand” or “A better theory is” with the use of a collaborative knowledge building tool. Schellens et al. (2007) proposed discussion structuring by assigning roles to each participant as “moderator, theoretician, summariser, source searcher” at the beginning of the discussion. Another approach was a forum with seminar leaders initiating the conversation while group members develop the argumentation accordingly and mediators intervene when required (Gerosa et al., 2010). Ligorio et al. (2005) conducted an experiment on collaborative writing of fairy tales by two distant primary schools’ kids in which, practitioners monitor pupil interventions during the writing and discussion stages, resolve conflicts in group work, summarize what pupils stated in discussion flows and create space for the pupils with less involvement within activity.

As shown in Oliveira et al. (2011), role of the online teacher/instructor varied, according to the activity phase objectives; for example, during first phase, teacher worked as a facilitator and a critical observer but during the group work stage the responsibility for the discussion leadership was entirely the responsibility of the participants while teacher is a passive observer only. Furthermore, academic moderators had been assigned in certain learning scenarios to lever the learners and also to ensure ethical etiquette and confidentiality in a discussion environment (Walta & Nicholas, 2103). In another article, instructors had introduced few online activities like quick-answer competitions or whole-class ratings as to spark collaborative online learning Zhan, Xu, and Ye (2011) and as in Ferriman (2013) and Ligorio et al. (2005), initial guidance, feedback or topic selection had been done by the teacher in the collaborative activity.

Most of the articles have used technology-mediated discussions such as discussion boards or online chat forums while a few had mentioned only about face-to-face communication. The hybrid of these techniques was also found where learners meet each other face to face as well as they meet in virtual environment (Samarawickrema, Benson, & Brack, 2010; Ladyshewsky, & Gardner, 2008). Another interesting observation is that a significant amount of articles had attempted both synchronous communications like online chat or video conferencing and asynchronous communication techniques like blogs, discussion forums within the experiment to implement collaborative activity (Samarawickrema et al., 2010; Brett, & Nagra, 2005; Wang, 2009; Raymond et al., 2005; Calvo et al., 2011).
Apart from the discussions, the other CL activities observed were peer assessments (Saunders, & Gale, 2011; Freeman, & McKenzie, 2002; Ligorio et al., 2005) with critical evaluation among peers to construct knowledge, collaborative writing spaces such as wikis (Calvo et al., 2011; Brett & Nagra, 2005; Oliveira et al., 2011; Li, Dong, & Huang, 2011) as a medium for socially mediated learning, or either game playing (Susaeta et al., 2010) or role playing (Ioannidou et al., 2010) depending on target activity and collaborative presenting of an end-result of a group collaboration (Tsai, 2010; Raymond et al., 2005).

In the case of the article portion that had participants greater than 100 as the sample size, open discussions surpasses (42.4%) and apart from discussions peer assessment (15.3%), a widely used CL pedagogical technique with larger learner communities, seems prominent too as indicated in Figure 4. More than 50% of the experiments had exercised combination of collaborative learning activities as explained by Oliveira et al. (2011) in the article, after studying and reflecting individually students participated in online discussions, small group work, collaborative writing spaces to share ideas and played games to find the final solution. Hybrids of virtual and physical discussions along with peer assessment are also (Zhan, Xu, & Ye, 2011) seen as fruitful to be implemented upon larger groups. Wang (2009) shows better collaborations using shared spaces with large classes implementing instructional design strategies such as friendship in groups and allowing meaningful tasks to improve individual accountability and positive interdependence along with certain scaffolding like collaborative writing of progress reports and monitoring groups individually.

![Figure 4. Percentage distribution of collaborative learning activities with experiment samples larger than 100 participants](image)

### Collaboration mechanisms

RQ4 refers on how to achieve potentially effective collaboration in a collaborative learning technological environment and the corresponding analysis code is “CL Structure” either mediated by the teacher or by learners themselves. Known collaboration mechanisms are group formation according to specific policies, or the progress along a learning flow, the distribution of roles and the distribution of knowledge / resources (Dillenbourg, 2002; Hernández-Leo et al., 2006). These mechanisms orchestrate the elements of a collaborative learning activity with the aim of triggering desired knowledge-intensive social interactions.

Group formation is the mechanism presented more in the analyzed papers. 40.88% of the papers discuss about a mechanism for group formation either by an external influence or individual willingness. Another 39.23% had mentioned a particular mechanism for distributing learning resources like learning materials or hardware devices within groups or any other component needed for the collaborative learning activity. Only 18.23% performed activity by assigning roles to group members whereas 1.66% had considered previous activities attempted by the student when grouping for collaborative activities.

Only very few instances were found conducting current learning activity by considering learner profile and previous performances (Capurro & Capretz, 2009; So, Seah & Toh-Heng, 2010). Research in many areas has shown that learning within groups improves students’ learning experience by enabling peers to learn from each other (Ounnas, Davis, & Millard, 2009). Hence, almost all papers had discussed forming groups either by students themselves or by the teacher. Forming groups by teacher or by a particular software program have been considered as designed-by-
teacher Group Formation (GF). This mechanism is used in a high quantity of papers (43.4%) as shown in Figure 5. Various practices were observed in literature for forming groups as GF according to geographical location or gender (Samarawickrema et al., 2010; Walta & Nicholas, 2013), using GF algorithms (Sancho-Thomas, Fuentes-Fernández, Fernández-Manjón, 2009), based on the type of activity, experience level or knowledge proficiency (Oliveira et al., 2011; Calvo et al., 2011; Tsai, 2010); but there was a considerable trend for random group formation (Shaw, 2013) and some articles had not expressed their mechanism.

Social resource distribution was encouraged in a collaborative environment when learners exchange resources over teacher mediation. However, in most cases the distribution meant a collective sharing of knowledge. Especially when forums or blogs were used, there was a tendency of exchanging extra learning materials relevant such as additional reading references in web or a link to a video along with the forum or blog post (Schellens, Van Keer, De Wever, & Valcke, 2007). The general trend of resource allocation was sharing learning materials in a LMS or another shared space to be accessible for all. A more truly fostering-collaboration mechanisms based on distribution of knowledge (e.g., to promote positive interdependence and individual accountability within groups) was found in situations of individual resource allocation when teacher allocates resources due to the requirements of the activity or a specific collaboration flow (e.g., Jigsaw activity) (Bochicchio & Longo, 2009; Hernández-Leo et al., 2006; Susaeta et al., 2010) or when students select the content according to their preferences (Wang, 2009).

Only 18.23% papers had discussed about assigning roles in a collaborative activity. The types of techniques for role allocation vary. While in Sancho-Thomas et al. (2009) role distribution was done by using a standard algorithm, in Schellens et al. (2007) the assignment is random. Besides, a negotiated role assignment between learners is also a valued approach since students can select their own roles according to their interests.

As shown by Figure 5(b), teacher/instructor involvement or providing certain scaffolding mechanisms for learner collaboration within the flow design was seen as a common practice even with larger learner communities (>100) rather than letting learners to decide independently. Zhan, Xu, and Ye (2011) had highlighted the significance of different grouping mechanisms and the effects on heterogeneous groups vs homogenous based on learner style in online learning environment as limitations within their experiment with larger learner crowd even though they only had used heterogeneous groups and teacher mediated resources and role allocation. Instructor presence has a heavy influence for the interactions during group work due to the supportive facilitator role as explained by Oliveira et al. (2011) even among massive learner communities.

Technological facets

Following figure 6 demonstrates usage of different technological facets and tools observed in the concerned literature and how past research had shown interest for seamless aspects. When considering different technologies suggested by the papers (RQ5), it was observed that many researches had used either the existing LMS at their institutions (e.g., Moodle, BlackBoard, WebCT etc.) or a similar platform to a LMS or customized the LMS according to the requirements and embed more tools like discussion boards (Shaw, 2013; Noroozi et al., 2013), conferencing or
communication tools (Williams et al., 2011), podcasting tools (Saunders, & Gale, 2011) and reflective journal logs (Ladyshewsky & Gardner, 2008), repositories with required learning materials (Zhan, Xu, & Ye, 2011) or even additional assessment or feedback mechanisms (Saunders & Gale, 2011) to make the final tool more sophisticated. Hence learners are able to experience a comprehensive learning environment with asynchronous or synchronous learning tools (Raymond et al., 2005). As indicated by figure 6(a), total 44.6% from the sample set had used either LMS or LMS-embed tools for their experiments. Another 30.9% had used generic tools such as Wikis, YouTube, Facebook, Blogs, Skype or Presentation tools like Microsoft PowerPoint in order to promote collaborative learning and aid in the learning activity flow. Some of the researches (24.5) had introduced completely new platforms and applications or new hardware devices according to their proposed conceptual model of the research (Ferriman, 2013). It can be derived that a significant number of research work oriented towards relatively high number of participants apply well-known LMSs, generic Web2.0 and communication tools, or LMS integrating specific tools.

Ubiquitous and pervasive computing offers new possibilities to work collaboratively mediating social interactions in technology-rich diverse spaces and times and across-technologies seamlessly. Novel experiences for learners including mobile technologies rises new opportunities for collaborative learning (Looi et al., 2010; Suo et al., 2009), also when seamlessly combined with other devices and software tools (Pérez-Sanagustín et al., 2012). Most of the research found in the search had not been able to achieve broad seamless features (figure 6(b)), even though certain attempts have discussed being seamless within their context. 57.4% of the papers have not considered being seamless across neither technologies nor space or time. 11.1% of the papers had achieved seamless in their solutions, mostly technologically and the other 31.5% discuss that being seamless is advantageous and effective. Also they had convinced productive activity flow is achievable across seams with their solutions.

It was observed that the suggestions for being seamless are augmented by combining several technologies together and offering them as a single solution (Tsai, 2010; Walta & Nicholas, 2013; Saunders, & Gale, 2011). In Calvo et al. (2011) cloud concepts are embedded to continue a collaborative writing activity for a group of learners in diverse contexts or conducting physical lab experiments in a remote virtual environment (Bochicchio & Longo, 2009). Noroozi et al. (2013) propose the use of multiple devices and technologies to conduct an activity involving face-to-face discussions as well as virtual discussion boards and other communication facilities extending the learning experience.

Discussion

The analyzed papers identify a number of challenges that arise in technology-supported collaborative learning environments involving a large number of learners. These challenges are related with technological, cultural and lingual barriers (Samarawickrema et al., 2010; Saunders, & Gale, 2011; Ferriman, 2013; Tsai, 2010; So, Seah, & Toh-Heng, 2010), and also with maintaining interest throughout the learning activity in order to promote students’ engagement and maximize the social interactions (Walta & Nicholas, 2013; Junco et al., 2013; Ferriman, 2013; Sancho-Thomas et al., 2009) that would lead to fruitful learning. Also as presented in Walta and Nicholas (2013), Wang (2009), Shaw (2013), Yang, Wang, Shen, and Han (2007) and Capuruço and Capretz (2009), minimizing the complexity of technologies and tools and making those flexible ensuring solid flow of learning activities, could be another challenge. Some papers had appreciated having a certain scaffolding structure that could also be mediated by
the teacher or instructor (So, Seah, & Toh-Heng, 2010; Wang, 2009). When considering the architectural models proposed, some attempts had reused an existing learning management system either by customizing it according to the needs or embedding new features/tools whereas others had practiced generic tools like Blogger, Skype, NetMeeting or social networking applications like Facebook or Twitter to promote learning. One frequent observation of the literature is that more than one technological approach had been utilized when designing the collaborative architecture as an alternative or complement to traditional LMS approach. Building learning communities by embedding more context-aware approaches such as artificial intelligent agents or contextualized knowledge bases were observed and those are seen as prominent and promising in future approaches (Calvo et al., 2011; Yang et al., 2007; Raymond et al., 2005).

Having homogenous student groups and small sample sizes for experiments had also been reflected as major concerns (Walta & Nicholas, 2013; Junco et al., 2013; Williams et al., 2011; Ferriman, 2013) by most of the researchers and had suggested that the experiments should have been extended to heterogeneous groups with larger sample sizes. This is interesting from a quantitative research perspective, but also important to evaluate the scalability of the technologies. While enabling communication is satisfactorily achieved by the current technologies, managing coordination and maintaining intense collaboration among many students are still challenging. When promoting collaborative activities like group formation, assigning roles or resource distribution teacher intervention had been seen as important by certain researchers. More articles recommended (Huang et al., 2010; Noroozi et al., 2013) forming groups as designed by teachers (manually or with software support), rather than letting students form their own groups according to their preferences since it would lead for subjective groups or free-passers because of friends. Adult guidance or collaboration with more capable peers (Vygotsky, 1978) in problem solving circumstances is encouraged in the learning sciences in order to achieve higher order cognitive levels. Learning being a continuous fluid process, when forming groups or distributing roles for activities, it would be an additional positive factor leading to potentially finest learning groups to be formed if learners’ track record could be considered. But it was revealed that only least effort had been taken in literature with this regard and future research in embedding student profile to orchestrate learning activities will be welcoming.

Conclusion

The literature review presented in this paper has offered an understanding of the research situation (till 2013) around collaborative learning technologies for arguably large numbers of participants. The quotient of learners considered for massive classes or large communities was in 55% of the top scientific journal papers selected under 100 participants, in 37% between 100-1000, and in 8% over 1000. The scenarios studied were predominantly contextualized in higher education settings, followed by primary and secondary education. Scenarios of adult education and informal learning are less common in the reviewed studies. Overall, mainstream software such as general Web2.0 tools and LMS are the main platforms being used in the studies; only a reduced number of cases considered pervasive technologies and specific tools devoted to support collaborative learning. Activities mostly based on communication actions (open or structured discussion) are widespread in massive scenarios as well as peer review tasks devoted to distribute the assessment workload between learners. However, activities requiring higher coordination between learners (e.g., collaborative writing) are rare. Similarly, group formation techniques of diverse type are considered in those activities. Other collaboration mechanisms like allocation of roles or knowledge distribution among group members to structure the collaborative activity are less frequent.

Pedagogical models and platforms for massive courses can benefit from existing research result in technology-supported collaborative learning environments (e.g., application of group formation techniques, considering different types of intervention by instructors). However, there are still important challenges to address in massive collaborative learning. These challenges are of different nature, from cultural to technological, but its core relies on being able foster knowledge-intensive social interactions. Some of the reviewed papers tackle this critical aspect, and there is an additional corpus of research (not considered in the review since it is) deliberately oriented towards small groups of learners, that propose a number of solutions to that purpose. Understanding how these contributions for small groups can be scaled up is an interesting future research line. A related, but different, perspective is the exploration of innovative collaborative learning approaches that may work better at massive scale (Ferguson & Sharples, 2014). An analysis of the different perspectives and building on top of this literature review will allow the formulation of CSCL design aspects for MOOCs. This paper provides a founding framework of conceptual aspects to be considered and a set of discussion pointers that lead to further research directions.
The paper has answered the question of whether research on collaborative learning technologies has addressed massiveness. The question is answered considering top established journal publications in educational technologies. Further literature review is focused on other fora, such as new publication venues centered on MOOCs or recent conference proceedings can also provide a complementary view – to the review presented in this paper – of how research and practice on collaborative learning technologies is now being applied at a massive scale.

Acknowledgements

This research has been partially funded by the Spanish Ministry of Economy and Competitiveness in the EEE Project (TIN2011-28308-C03-03).

References


Facebook Groups as an Academic Teaching Aid: Case Study and Recommendations for Educators

Eli Miron* and Gilad Ravid

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer-Sheva, Israel // emiron@post.bgu.ac.il // rgilad@bgu.ac.il

*Corresponding author

(Submitted August 20, 2014; Revised December 4, 2014; Accepted April 22, 2015)

ABSTRACT

The move from a walled garden type Learning Management Systems (LMS) to open environments (like Facebook) forces us to adapt new teaching ways. This article offers a brief review of the use of Facebook groups in learning, describes the experience of using Facebook groups in an academic institute, explains the considerations for choosing the type of group and provides detailed technical guidance for teachers and students, including recommendations for enhanced privacy and Internet security and for reduction of information overload. Good technical understanding of the Facebook platform and active participation (at least once daily), are recommended for successful use. Facebook groups have been used by the authors during 2012 – 2014 in 12 courses. Overall results show that the use of Facebook groups for academic purposes is favoured by the students. From the educators’ vantage point: communication with the students was fast and easy. Email alerts enabling communication with the students, but without the need to “live in Facebook” an answer to one student was visible to all. To sum up: the experience of both the students and the authors is favourable. Finally, the acceptance of Facebook as an LMS was analysed using a simplified version of the Technology Acceptance Model.

Keywords

Social network sites, Blended learning, Privacy, Learning management systems

Introduction

The use of social networks in teaching is not only a matter of convenience. In the introduction to his famous article Digital Natives, Digital Immigrants, Marc Prensky (2001) noted that “Our students have changed radically. Today’s students are no longer the people our educational system was designed to teach” (bold typeface in the original). Furthermore, he stated that “because the single biggest problem facing education today is that our Digital Immigrant instructors, who speak an outdated language (that of the pre-digital age), are struggling to teach a population that speaks an entirely new language.” Four years later Prensky (2005) suggested that “Students want and deserve to receive this content through 21st century tools that are powerful, programmable, and customizable —through tools that belong to them. We could offer this content to them on their cell phones, for example.”

Integrating networking technology into classes can be done in many ways, on the continuum of no technology at all at one end, to fully-online courses on the other end. In between there are various options to conduct a blended learning course (Garrison & Kanuka, 2004; Graham, 2006; Vadiathan, 2002). Most of the implementations of technology enhanced learning (TEL) use a walled garden solutions, mainly a LMS systems (Alier, Casañ, & Piguillem, 2010; Bhattacharya & Dron, 2007). Those systems protect the learners from outer content, supply the adequate intimacy needed for social presence (Noy, Raban & Ravid, 2006) and provide a sense of a learning group that works together in the learning process. At the same time, walled garden solutions force the learners to move from the platforms which they use for their daily computerized activities to the propriety systems. Recently, educators started replacing LMS systems with open systems. Open systems have many advantages and can be integrated easily with the learner regular activities and make him learn all the time. While open systems have many advantages, we are still forced to solve some challenges, for example how to supply the group sense to each learner. In the early days of using Internet technology in classes, educators complained that they need to maintain the course site and answer to student’s e-mail 24/7. We hope that the use of open systems, like Facebook, for learning will enhance their use outside regular teaching hours and extend the class to the external environment.

Facebook (FB) is one of the “21st century tools” that are the playground of the digital natives and can be used for teaching purposes. By August 2013 the total number of Facebook (FB) users has reached 1.5 billion!! (Smith, 2013).
The percentage of Facebook users in the age group of 18 – 29 (the age of most students) approaches 90% (Brenner & Smith, 2013). This means that a very high percentage of university and college students are very familiar with Facebook. Furthermore, they login to Facebook a few times daily and a large fraction of them practically “live” with Facebook. The familiarity of the Facebook platform and its very frequent use by students makes it an appealing candidate for various learning enhancing applications, in parallel, or instead of other university platforms such as “High Learn” (the Israeli version of Fox system by Britannica), Blackboard, Moodle or others. However, being a social network site, Facebook is open by design. This open nature of Facebook prevented its use in classes since it could not provide adequate intimacy. The Introduction of Facebook closed and secret groups at 2010 enabled the creation of closed, course-specific communities, where only the lecturer, instructor and students constitute the group members. These self-contained communities can be used for asynchronous and synchronous interactions in an academic course. The use of Facebook groups also allows sharing of information, documents, pictures, links to websites, etc. The open nature of the Facebook group (to its members only) provides a convenient platform for cooperative and/or collaborative learning.

The difference between collaborative learning and cooperative learning was explained by Panitz (1996, 1999): Collaborative learning is a personal philosophy, not just a classroom technique. It suggests a way of dealing with people which respects and highlights individual group members’ abilities and contributions. There is a sharing of authority and acceptance of responsibility among group members for the group’s actions. Cooperative learning is defined by a set of processes which help people interact together in order to accomplish a specific goal or develop an end product, which is usually content specific. Cooperative system is more directive than a collaborative system of governance and closely controlled by the teachers.

Since Facebook groups are essential for the use of Facebook in classrooms, we start by ranking the various types of Facebook groups by their privacy options, then list various uses of Facebook groups for academic teaching, describe the experience of using Facebook groups in a University setting and conclude by providing detailed “how to” recommendations for lecturers and students with an emphasis on privacy and Internet security settings.

Practical use of Facebook as a teaching aid in academic courses

Recent publications discuss the use of Facebook not only as a means of communication between professors to students, but for various other academic uses. Fordham and Goddard (2013), list various application modes of Facebook in teaching and learning, divided into three categories: “formal learning,” “non-formal and out of school hours learning” and “wider applications.” The use of Facebook in formal learning consists of the following:

- Creating a Time-line or Facebook Group to support the teaching of any curriculum subject.
- Creating a space and platform for homework and revision resources.
- Running debates on topical issues and hot issues in the media.
- Peer tutoring and support.
- A research tool to post, share ideas, videos and resources.

Meishar-Tal, Kurtz, and Pieterse (2012), discuss the use of Facebook as a LMS: “The role of the LMS is to serve as a platform for course sites and to fulfill three goals: (1) to provide students with digital learning materials, such as articles, presentations, summaries of lessons, and arrange them in a way that reflects the course plan; (2) to employ interactive learning activities with students in the forums, wikis, and other collaborative tools; and (3) to manage the course and the learners, maintaining tests, evaluating the students’ learning and achievements, and giving grades online.” For brevity and further reference we denote the three goals as LMS -1, LMS-2, and LMS-3 respectively. The authors conclude that “the case study described above demonstrates that design and operation of a learning activity within a Facebook group produces a very intensive and collaborative learning process.” Similar uses have been described by others as well (Fardoun, Zafar & Ciprès, 2013).

Wang, Woo, Quek, Yang and Liu (2012) discuss their experience of using Facebook’s closed groups as an academic learning management system. Their conclusions are: (1) “The finding of this study confirms that the Facebook group has the potential to be used as an LMS.” (2) There are constraints resulting from safety and privacy issues: “In this study, the Facebook group was set to “closed,” and the students were not required to be friends. The students, in particular the Master students, however, still did not perceive it as a safe environment. They were commonly worried about, on the one hand, their academic performance in the course could be discovered by their social friends; on the
other hand, their personal information and social lives might be accessed by the tutor. This study confirms that privacy and Internet safety become a critical concern in social learning environments.” (our emphasis), but the authors do not suggest how to overcome these safety and privacy concerns. We will elaborate on these concerns in the proceeding paragraphs, and describe in detail some practical recommendations how to overcome this critical concern.

In reality, traditional LMS systems also have privacy issues that are assumed risks with class enrollment. Students know each other’s names and many of the introduction activities may include optional shared personal information. Face-to-Face classes usually include introductions, also. Facebook as a companion to an LMS should be aware of privacy concerns but not to an extreme.

**Privacy of Facebook group types**

Facebook allows three types of groups: open, closed and secret, the differences among them is the degree of privacy. The different privacy options are listed in Table 1 (Facebook Inc., n.d.):

<table>
<thead>
<tr>
<th>Types of Facebook groups</th>
<th>Open</th>
<th>Closed</th>
<th>Secret</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who can join?</td>
<td>Anyone can join or be added by a member</td>
<td>Anyone can ask to join or be added</td>
<td>Anyone, but they have to be added</td>
</tr>
<tr>
<td>Who can see the group name and who’s in it?</td>
<td>Anyone</td>
<td>Anyone</td>
<td>Only members</td>
</tr>
<tr>
<td>Who can see the group description?</td>
<td>Anyone</td>
<td>Anyone</td>
<td>Only members</td>
</tr>
<tr>
<td>Who can see the group tags?</td>
<td>Anyone</td>
<td>Anyone</td>
<td>Only members</td>
</tr>
<tr>
<td>Who can see what members post in the group?</td>
<td>Anyone</td>
<td>Only members</td>
<td>Only members</td>
</tr>
<tr>
<td>Who can find the group in search?</td>
<td>Anyone</td>
<td>Anyone</td>
<td>Only members</td>
</tr>
<tr>
<td>Who can see stories about the group on Facebook (like in News Feed and search)?</td>
<td>Anyone</td>
<td>Anyone</td>
<td>Only members</td>
</tr>
</tbody>
</table>

It should be noted that in all group types, each member of a group can navigate to the time-line (wall) of any other member, and see his activities and his friends’ activities. This transparency produces a mixing of the learning environment with the private environment of the group members. Detailed instruction how to prevent such mixing are given in Appendix A. Open groups can be discovered by regular Internet search. However, search engines such as Google do not index closed and secret groups and therefore these Facebook groups cannot be found by search outside Facebook (i.e., by using Google). However, closed groups (not secret groups) can be found in a search within Facebook. It should be noted that anyone who finds a closed group can see all its members - names and pictures (sometimes even job and place of residence), but would not see members who have undertaken specific precautions – which are detailed later. In addition, anyone who locates a closed group in a Facebook search can apply for membership – many such applications could be a nuisance.

Only secret groups provide a complete “members only” environment. However, it requires more work to establish - new members have to be personally invited. Facebook friends can be added directly by the group administrator. However, when an educator opens a secret group, the students have to be invited using their email addresses, since usually, students are not Facebook friends of the educator or the instructor.

**Academic use of Facebook groups at an academic institution – case study**

Facebook groups have been used for communication with students in twelve knowledge management courses held during 2012-2014 at an academic institution. It should be emphasized that all other courses that the students attended were managed conventionally by a propriety LMS (HighLearn or Moodle). Next we examine how the Facebook groups that we used met the objectives of the learning management system as mentioned above.
Communication educators-students

The Facebook groups were used at first mainly for organizational communications between the lecturer/instructor and students such as clarification on submission dates of assignments, posting lists of scores (by personal code, not by name) and more. The Facebook group provides an easy way of recognizing the student (helps in showing familiarity and adding a personal touch): clicking the “about” tab (circled in Figure 1), which displays the pictures of all group members who have a picture in their Facebook profile (Figure 2).

![Figure 1. Facebook group main toolbar](image1.png)

The display of the pictures of all group members is active for the group creator only.

![Figure 2. Group’s members](image2.png)

This Facebook page has been anonymized (random names) for privacy using Social Fixer (see Appendix B).

Clicking the down arrow near “All members” (circled) provides sorting options. Clicking “Message Members” (squared) shows a multi-chat screen where chat partners can be chosen.

Sending a personal message to one of the group members by any other member is enabled by clicking on the member’s name.

The list of students’ pictures helps the educator to recognize the students and connect names to faces – thus adding a personal touch to the classroom. Even though we have 100% participation (mandatory), not all students appeared on Facebook with their pictures (Table 2) that includes data for B.Sc. and M.Sc. students. Most M.Sc. students study only part time. Some were working at sensitive workplaces which restricted the use of Facebook - hence the relative large number of blank pictures. In other courses most students did display their picture.

<table>
<thead>
<tr>
<th>Students type</th>
<th>Total Number of students in course</th>
<th>Number of students with profile picture</th>
<th>Percent of students with profile picture</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.Sc.</td>
<td>39</td>
<td>39</td>
<td>100</td>
</tr>
<tr>
<td>B.Sc.</td>
<td>34</td>
<td>33</td>
<td>97</td>
</tr>
<tr>
<td>M.Sc.</td>
<td>29</td>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td>B.Sc.</td>
<td>33</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td>M.Sc.</td>
<td>30</td>
<td>29</td>
<td>97</td>
</tr>
<tr>
<td>B.Sc.</td>
<td>39</td>
<td>39</td>
<td>100</td>
</tr>
<tr>
<td>B.Sc.</td>
<td>10</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>M.Sc.</td>
<td>20</td>
<td>17</td>
<td>85</td>
</tr>
<tr>
<td>M.Sc.</td>
<td>26</td>
<td>19</td>
<td>73</td>
</tr>
<tr>
<td>MBA</td>
<td>55</td>
<td>37</td>
<td>67</td>
</tr>
<tr>
<td>M.Sc.</td>
<td>34</td>
<td>29</td>
<td>85</td>
</tr>
<tr>
<td>B.Sc.</td>
<td>49</td>
<td>49</td>
<td>100</td>
</tr>
</tbody>
</table>
Event scheduling

Facebook allows for organization of events (real and virtual) such as irregular meetings, or schedule additional exercises. The events are created using the events tab and “+ Create event” (Figure 3).

Surveys and polls

Using the built-in option “Ask a Question.” Facebook allows the preparation of a structured questionnaire (Figure 4), with answers can be prepared in advance to choose from. The students found this option convenient for their purposes and used it for coordination among them. The following example (Figure 4) was modified to show the options in English (original post in Hebrew).

Conclusion: all uses listed above are consistent with LMS-3.

Links to Students’ homework in the academic institution Knowledge Management courses include some online assignments: personal blogs on the subject of lectures, wiki group exercise (article summary or KM case study), YouTube videos clips on the subject of knowledge management and mashup creation. The students were required to post links to their online work in the Facebook group. Any student’s activity summary is easily obtained by performing a search on a student’s name in using the internal group search option (magnifying glass icon on the group heading line) which provides a condensed list of the all contributions of the student (LMS-2).

Cooperative learning

The existence of links to the online assignments provided easy access to personal works of other students and so that they were able to comment on them (a part of the assignment in some courses). The lecturer also posted some problems connected to the learned material and the students were asked to suggest solutions (non-obligatory). The
problems and comments initiated quite interesting discussions. Sample questions were posted on the Facebook group to help students prepare for the final exams and the lecturer responded to students’ comments regarding these problems. The students also used the Facebook group as a public forum to ask the questions about the forthcoming test. The questions, answers and following discussions were available to all students. These examples are indications to the potential of Facebook groups for cooperative learning. The Facebook group page used as a platform for students to post links to recent publications (updated or appeared two weeks before) relevant to knowledge management. Posting of these “news articles” was not obligatory. The incentive to participate was a simple reward system: every student that published a relevant knowledge item was enrolled as a participant in a books lottery held in the final lecture’ maximum one entry per student per week. An example to a news item is shown in Figure 5.

![Figure 5. Students news post (example)](image)

**Uploading course material**

Lecture files were directly uploaded to the Facebook group in cases where the number of relevant files was small (number of files ≤ number of lectures). The advantage of direct uploading is having everything on the same platform. One disadvantage is that the uploaded files are dispersed over the time-line. Searching for uploaded files in groups is easier using the “files” tab (circled in Figure 6).

![Figure 6. Facebook’s files list display button](image)

Clicking the “files” tab displays only the list of uploaded files, in chronological order (the latest is the uppermost), and thus locating files is much easier.

An interesting option in Facebook groups is creating a new file directly from Facebook. It is similar to using Google Docs but provides fewer options. To create a new document: press “+ Create Doc,” name the document (and write the contents) and then press “+ Create Doc” again (Figure 7).

![Figure 7. Create a file within Facebook](image)

In cases where we have many files we chose Google Drive as the platform for providing course material to the students, in a similar way to Meishar-Tal et al. (2012). The educator opens a course folder in Google Drive and shares the link with the students by publishing it in Facebook. All files relevant to a lecture were uploaded to a corresponding sub-folder. All new files uploaded to Google Drive are immediately available to the students. The
sharing option used by us was that everyone that has the link could read (and download), not edit. It is sufficient to put the link in Facebook. We pin the message with the Google Drive link is pinned to the top for easy access. Meishar-Tal et al. (2012) recommend to write the link to the course’s folder (on Google Drive) in a file and post the file in the Facebook group. The file is easily accessible using the file tab, preventing the need to look for the link downstream. All these options comply with LMS-1 requirements. Another advantage of using Google Drive is the availability of built-in sophisticated questionnaires (Google Forms). It is easy to write a questionnaire, put it in the course folder and the results are automatically summarized in an Excel file.

**Basic requirements for using Facebook groups in class**

- The teacher, usually a “digital immigrant,” must be proficient in the use of Facebook.
- All students in the class must use it (be group members).
- The Facebook group should legitimately co-exist with current other Learning Management Systems in the university, e.g., Moodle.

**Challenges in using Facebook groups for academic teaching**

**The practical ability of students to use many Facebook groups simultaneously**

As far as we know, this issue has not been tested for Facebook groups that were used as an LMS of some sort. In some academic institutes students opened a few Facebook groups for interaction among themselves in course related issues. For example, first year engineering students in another academic institute in Israel opened themselves 5 closed Facebook groups for discussions – each group deals with a single major course. Our students used a Facebook group for all the discussions between themselves. In addition, students are usually active in many other Facebook groups – not related to learning. Although not a proof, it can be viewed as an indication to the ability of students to handle multiple Facebook groups. Unwanted notifications (noise) can be reduced by using the option to filter notifications, or by turning them off from the “notifications” menu in the top line of the group page. Another, more sophisticated option, is to use Social Fixer’s “advanced filtering” (see Social Fixer, Appendix B).

**Co-existence of Facebook groups with other teaching platforms**

We received a few complaints from students that it is not convenient to use multiple LMS environments simultaneously, but most students did not observe any difficulty.

**Privacy and internet security issues**

We experienced concerns and objections, similar to those that have been described by Wang et al. (2012). Some students, especially master’s students, are not digital natives, and do not behave as such, especially with regard to privacy. A recent research carried on 441 subjects (Kuo & Tang, 2013) found that “research has indicated personality is one of many factors may have some influence on Facebook's usage, information disclosure.” We experienced some students who were “Facebook illiterate” and some who were paranoid. Even though these were a minority, we had to find solutions how to get them involved in, since participation was a requirement. We are not aware of solutions to privacy issues pertinent to Facebook groups in education that have been previously published. Security and privacy issues are sometimes an obstacle that must be taken out of the way prior to the implementation of Facebook groups in education. Possible solutions and further describe them in detail in Appendix B. The very few students that would not use Facebook groups due to privacy concerns, even with the measures we describe, were allowed to use Facebook anonymously (requires opening an anonymous email address, e.g., xyz123@gmail.com). The only single strict requirement was that students must provide their anonymous name for grading purposes.
What students think about using Facebook

We asked a few past students about their personal experience with the use of Facebook groups in class. Overall, the majority of responses were positive or even very positive, for example: “I felt much more comfortable on Facebook than with other LMS, one push of a button is enough to have everything appear and open to everyone.” They mentioned that most courses are using the other LMS so they have to handle and use many platforms, which caused a problem of multiple platforms. The students would prefer that all courses will be conducted in the same platform, namely Facebook. This raises a question: will the need to handle several Facebook groups simultaneously be convenient for students? Further research is needed to provide the answer to this question. In addition we can report that during the course some students took advantage of the Facebook platform for issues unrelated to learning (Figure 8).

Figure 8. Non course-related post

The translation of the post (from Hebrew) - XYZ company is hiring: looking for students who graduated in 2011 or 2012 and are interested in a marketing career. Those interested should contact me (the student) via Facebook.

A detailed questionnaire regarding the use of Facebook in academic education was distributed among students of some courses, 77 students responded. Most students (49) were at the age of 26-30, 21 were younger and 7 were older. The students reported on the frequency of posting in the course Facebook group: 3 students (4%) published on average more than one post per week, 23 (30%) published on average 1 post per week and 51 (66%) were just passive participants. We regard active participation of the about a third of the students as a success to the use of the Facebook platform. 51 students regard the Facebook group as a convenient platform for communication with fellow students, however only 38 (49%) view this as a convenient platform for communication with the professor and/or the teaching assistant. We have no explanation to this difference since the Facebook group was almost the single channel for such communications.

About half the students (38) regard privacy issues as a problem, 31 of them belong to the elder students. The use of closed or secret Facebook groups resolves most privacy concerns. It is interesting to show the detailed response to the question “do you think that a Facebook group is more convenient than other LMS in the university?” 15 students (19%) gave a grading of 1 out of 5, 8 students (10%) graded it as 2, 19 students (25%) grade it as 3, 17 students (22%) gave 4 and 18 students (23%) gave 5. These results show that altogether the use of Facebook groups for academic purposes is favored by the students. We find high correlation between the score the students gave for the convenience to communicate with the lecturer and the convenience to communicate with other students ($r = 0.57$) and preference to use it in academic settings ($r = 0.62$). There is high correlation between the convenience to communicate with other students and the importance to discuss various subjects ($r = 0.5$). Students who were concerned with the ability to watch their time-line were also concerned with the susceptibility to Internet attacks ($r = 0.6$). The only difference between males and females is with the number of times they check their Facebook account daily: males check it 2.88 a day while females do it 3.53 times. Comparing the different age group reveal that the only differences are in number of times they check their account and post to it.
From this plot (Figure 9), it is clear that students older than 31 visit Facebook less frequently.

![Boxplot of question “number of times checked Facebook account”](image1.png)

*Figure 9. Boxplot of question “number of times checked Facebook account”*

From this plot (Figure 10), it is clear that students of the youngest and eldest students are those that post more frequently in the class Facebook group.

![Boxplot of question 6 “How frequently do you post to the course Facebook group”](image2.png)

*Figure 10. Boxplot of question 6 “How frequently do you post to the course Facebook group”*

**Analysis of the usefulness of Facebook groups as an LMS using the Technology Acceptance Model (TAM)**

Davis (1989), suggested a Technology Acceptance Model (TAM) based on the insight that users’ attitudes toward technology are critical factors in their accepting and using new technologies. Perceived usefulness and perceived ease of use are the most fundamental determinants for formulating positive attitudes toward technology and behavioral intention to use technology and, therefore, ultimately define actual use. The acceptance of Facebook as a social network has been analyzed by Rauniar, Rawski, Yang and Johnson (2014), using TAM. We analyze the
acceptance of Facebook as an LMS using a simplified version of the Venkatesh, and Davis (1996) “final model” version of the Technology Acceptance Model for the following reasons (Figure 11):

(A) The use of an LMS in an educational institution is imposed - the students are not given a choice. (B) there are two actual user groups: students and educators (lecturers and instructors). (C) Facebook is already been in use by the students and does not constitute a dedicated learning platform.

Facebook (coupled with Google Drive) is not a dedicated LMS and has many other uses. Both authors, who are savvy Facebook users, were the only lecturers who used Facebook for teaching in the university. The two course instructors used Facebook without any difficulty; hence the research was performed on students only.

Perceived Ease of use: following Zipf (1949) we define perceived ease (PE) of use as the least effort, i.e., each individual will adopt a course of action that will involve the least average work. In this respect Facebook has a great advantage – the students are already there: no new platform to install, no new technique and/or procedure to learn - practically zero learning curve. Furthermore, they enter the platform on daily basis for other purposes - no need for any specific login procedure and additional username/password combination. Most of the students reported that they are familiar with Facebook and know how to operate and handle the environment. However, the co-existence of other LMS for all other courses reduced the perceived ease of use, because of the need to use two environments side by side. Future use of the Facebooks groups as a single environment has a potential to eliminate this difficulty.

Perceived Usefulness: Perceived Usefulness (PU) is a user’s assessment of his/her “subjective probability that using a specific application system will increase his or her job performance within an organizational context” (Davis 1989). The educators regarded the use of the Facebook as LMS as “providing the goods.” From students comments we learned that the usefulness of Facebook as LMS is a double edged sword. On one hand, they are frequently on Facebook for other activities hence they can interact with their classmates and the study materials (some reported that the blending of hedonic activities and learning ones is undesired for them). On the other hand, the traditional LMS tools incorporate a better user interface for learning, most notifiable comment was that important information and knowledge “scroll away” from their attention too quickly.

Actual Use: Since users (students) have a high perceived ease of use and high perceived usefulness, the introduction on Facebook as a LMS was adopted easily and favored by the students. Another perceived advantage was the promptness of the educators’ responses, and those were viewed by almost all students. The class group was also used by the students for easy communication of messages for non-class purposes. Students also opened students-only Facebook groups for collaborative learning of class material for various classes, even those that used traditional LMS.

Conclusion

This paper reports the practical use of closed and secret Facebook groups in 12 knowledge management classes in a university demonstrating the feasibility and advantages of Facebook groups (augmented by Google Drive) as a learning Management System. It should be noted that Facebook can be conveniently used on various platforms such as tablets and smart phones, in contest with dedicated LMS platforms, such as Moodle, where the smartphone applications are quite inconvenient.
Qualitative analysis using the Technology Acceptance Model shows the reasons for the potential of using Facebook as an LMS. Due to its convenience students used this platform for both learning-related and non-learning-related purposes. Further research is required to substantiate this claim for general institutional use. The article provides unique, very detailed, how-to guides for implementation of Facebook groups in high education, including specific instructions for overcoming privacy and Internet security concerns - existence of scientific articles with such detail is unknown to us. This article can therefore be used as a practical implementation guide for educators who wish to utilize Facebook groups as a learning platform in academic institutions.

References


Appendix A

Tips for opening and using Facebook educational groups

Creating a group

Clicking on the Create group in the left column of the home page displays the Facebook privacy options of the new group. The first step is entering the group name (obviously related to the course). The second step is to choose the privacy level (closed or secret) and press create. The next step is to add at least one additional group member (a new group must include at least one member in addition to the group creator). This first member can be selected only from the group creator’s friends list. If the instructor is not a member of the educator’s personal Facebook friends, a random member of the group creator’s Facebook friends can be added (and deleted from the group after at least one student joined). With closed groups the students are given the group’s URL or they can find the group name in Facebook search and then apply for membership. With secret groups this option is not available. New members must be added from the list of Facebook friends, or invited by mail. It is therefore recommended to assign the instructor (or the educator) as an additional group administrator. For opening secret groups it may be convenient to open it at first as a closed group (easier to add members) and then convert to group to a secret one.

Appendix B

Privacy and Internet security

Imaginary privacy problems: New content published within a group is displayed on the home page of any personal Facebook account. This is actually the news feed, and it is visible only to the Facebook account owner himself. This content does not appear on the “time-line” (wall) which all his Facebook friends can see. Some students were not aware of this difference and complained that their Facebook friends (not only the group members) can see all the group content. Therefore it is important to explain this difference.

Privacy and safety measures: Educators and students concerned about the privacy and Internet security can add one or all of the following four “calming elements” which are not extreme and easy to add:

- Prevention of the ability of a group member (in our case educator and/or student) to see the “real” Facebook friends of another group member and enter their Facebook pages. This setting is important in general with regard to internet safety (Kruse, 2013). The default setting of Facebook is that friends’ lists are open to the public. To change these settings: go the personal time-line and click on “friends” and choose “edit privacy” (Figure 12).
Here one can choose who can see his friends list and which people and lists one follows. The most stringent setting is “only me.” The less stringent privacy setting, but still helpful, is “friends of friends.” This specific setting is not visible at the main menu and can be accessed only as a sub-menu of “custom.” This setting however is enough to prevent students/educators mixing.

More privacy options: go to the gear wheel in the Facebook home page of your account and choose “privacy settings.” An extensive description of Facebook privacy settings that has recently been compiled by Elliott (2013) and can be used to choose additional appropriate privacy settings.

Facebook specific free antivirus - Bit Defender SafeGo. SafeGo provides additional protection against scams, spam, malware and phishing attacks. SafeGo homepage is: http://bit.ly/1esW6UH . SafeGo installs as a Facebook application.

Forcing encrypted surfing. Facebook can be set to use HTTP Secure browsing (HTTPS) protocol, in which the browser uses an additional encryption layer to protect the traffic. Facebook calls this type of browsing “secure surfing.” Setting secure surfing: click on the gear in the top right side of the home page, and then click “Account Settings.” Click on “security” at the left top of the page – the uppermost option is “secure browsing.” If current setting is disabled, click on edit to change.

Social Fixer browser plug-in (extension). This plug-in allows to easily and comfortably manage all privacy settings on Facebook. Another important bonus is convenience of writing comments on Facebook. Usually using the enter button while responding or commenting in Facebook, uploads the comment instead of creating a new line. Social Fixer has an option to automatically require using tab + enter combination to upload (Figure 13).

A convenient Social Fixer option is to add more control on posts, such as “mark read.” When this option is enabled, hovering the mouse near the upper right side of the post displays 5 more option icons (e.g., mark read), as shown in Figure 14:

Another enhancement to reduce information overload is advanced content filtering: “Feed Filters give you full control over what stories you want to see or hide. It also lets you define rules to move stories to separate tabs so they are logically grouped how you want them to be.” Figure 15 shows two examples: The first is a filter that hides all notifications from the friend who is second from the top in the friends list (names are hidden for privacy). The second filter eliminates all content containing the word sex.
We recommend reading (Gordon, 2012 and Kruse, n.d.) for an extensive review on Social Fixer. 

*Note.* Social Fixer is available for all common browsers except Internet Explorer.
Designing a Resource Evolution Support System for Open Knowledge Communities

Xianmin Yang1 and Shengquan Yu2*

1Institute of Education, Jiangsu Normal University, China // 2Institute of Modern Educational Technology, Beijing Normal University, China // yangxianmin8888@163.com // yusq@bnu.edu.cn

*Corresponding author

(Submitted August 24, 2014; Revised December 12, 2014; Accepted March 4, 2015)

ABSTRACT
The continuous generation and evolution of digital learning resources is important for promoting open learning and meeting the personalized needs of learners. In the Web 2.0 era, open and collaborative authoring is becoming a popular method by which to create vast personalized learning resources in open knowledge communities (OKCs). However, the essence of openness of OKCs also gives rise to concerns regarding the knowledge quality and non-orderliness of resource evolution. In this study, we design a resource evolution support system (RESS) called learning cell system (LCS) in one OKC. Two key issues, namely, the intelligent control of content evolution and the dynamic semantic associations between resources, are addressed by combining technologies of semantics, trust evaluation, rule-based reasoning, and association rule mining. One typical case is taken to illustrate the actual evolution process of learning resources assisted by RESS in LCS. The operating effect of this system shows that RESS can control content evolution and effectively build semantic associations among resources. Finally, the academic contribution to the OKCs, implications for educational practice, limitations, and future research plans are presented.

Keywords
Open knowledge community, Learning resource, Resource evolution, Evolution control, Semantic association

Introduction
With the spread of Web 2.0 ideas and technologies, open knowledge communities (OKCs) have become increasingly popular. OKCs can be used as knowledge management tools and virtual learning environments for learners. Considerable research has recently been conducted to examine the potential of OKCs in educational application scenarios. For example, some scholars investigated the effect of student-generated content (SGC) on student engagement and learning outcome in wiki-type OKCs (Li, Dong, & Huang, 2011; Wheeler et al., 2008). In OKCs, any valid user can create new learning resources and coedit existing resource content (Yang et al., 2014). The involvement of many users ensures the continuous generation and evolution of learning resources while meeting the growing personalized needs of learners. Nevertheless, a credibility issue is associated with this collaborative authoring approach (Luo & Fu, 2008). Inevitably, a large number of non-credible, poor-quality contributions exist in this type of open and unmonitored environment. These contributions hinder the effective and efficient use of OKCs in the open learning era.

The complexity of the user groups and the liberalization of production in OKCs directly result in the explosive growth and disorderly evolution of digital resources. For example, facing the pressure of trust crisis (Seigenthaler, 2005; de Laat, 2014; Sapienza & Zingales, 2012), Wikipedia is setting up an increasing number of rules, including administrative rules, fact-checking rules, real-time peer review rules, and rules that prohibit the creation of new articles by anonymous users. In addition to these rules, Wikipedia also provides manual revision and check functions based on feedback. The purpose of all these measures is to ensure the high quality and orderliness of resource evolution (Wang, 2009). How to control the content evolution of learning resources adequately has become a realistic problem that should be addressed (Yang, 2012).

In addition, resources should not be isolated from one another to promote effective and efficient learning. Along with the development of semantic web technologies, the establishment of semantic associations among learning resources brings significant value according to connectivism theory (Siemens, 2005). First, the resource retrieval burden of learners is reduced, and target resources can be efficiently located. Second, an individual resource associated with many external entities will have a higher probability of being retrieved, viewed, and improved. Currently, the
semantic association among e-learning resources in OKCs is insufficient. Moreover, most associations are described by static metadata and built by manually editing hyperlinks between resources (Shao et al., 2008; Yang et al., 2013), such as the entity associations in Wikipedia and Hudong Baike (http://www.hudong.com). The integration of semantic technology and resource association has the potential to address the aforementioned problems, which can aid in the development of OKCs and open learning.

Literature review

Definition of resource evolution

Ecology is a discipline that focuses on species distribution, diversity, interactions among and between species, as well as the external environment (Begon et al., 2006). An integral ecosystem has the features of entirety, openness, dynamic equilibrium, self-organization, and sustainable evolution. The concept of evolution stems from ecology. Evolution is the change in the inherited characteristics of biological populations over successive generations. The application of evolutionary principles has occurred in different fields, such as economics, culture, computer science, and management. Evolution refers to the process of developing from simple to complex (Zhao, 2006). Development, change, and adaptation are the core ideas of evolution. Organisms change and grow gradually by interacting with and consequently adapting to their external environment. The evolution of learning resources specifically refers to the improvement and adjustment of resource content and structure (Yang & Yu, 2011). The aim is to meet the dynamic and personalized learning needs of learners. The orderly control of evolution means that an individual resource (or an entire resource population) develops in a desired direction in a quality-controlled manner.

Models of resource evolution

Resource evolution may be described according to two different models: content evolution and association evolution.

Content evolution

Content evolution refers to the constant updating and improvement of resource content through the collaborative content editing of many users (Yang & Yu, 2011). The process of content evolution is exhibited by continual content version changes.

Figure 1. Content evolution of learning resources

Figure 1 shows the process of content evolution. The producer creates and publishes a learning resource and attracts collaborators to edit the content. Given that the resource is available to the public, any user can revise content, add photographs, or insert external RSS (Really Simple Syndication) resources. As the resource grows, more users come into contact with it. These users comment, add activities, and annotate entries. By harnessing this collective wisdom, the content is constantly updated. Finally, the content evolves into a high-quality learning resource that meets the learning demands of different users.
Orderly control of resource content is a core issue in the evolutionary study of learning resources. In an open environment, resource evolution is centered on Web 2.0 technologies. For quality control, manual checking is commonly used to ensure that evolution is appropriate. With the help of manual labor, the versions evolve toward a higher quality and reliability. The quality control model based on manual checking is time- and labor-intensive. Ultimately, the quality control model fails to meet the requirements of dynamic generation and infinite expansion of learning resources.

**Association evolution**

Association evolution is the process of establishing and enriching semantic associations, which are mainly exhibited by the expansion of the resource network (Yang & Yu, 2011).

![Figure 2. Association evolution of learning resources. Note. R, resource; RG, resource group.](image)

Figure 2 shows the ideal process of association evolution. From the time of its inception, an individual resource strives to connect with other resources to enhance survival capability. The resource establishes an increasing network of semantic associations with other resources through manual operation or automatic discovery. Semantic relationships may include similarity, hyponymy, precursor, or equivalence. The growth of these associations provides data support for resource aggregation. With the use of specific resource aggregation mechanisms, resources that are similar will automatically form subject-based resource groups, whereas resources in semantic order will form knowledge chains with different learning priorities. Finally, any learning resource will become one node in an infinitely expanding resource network and will subsequently develop through dynamic association with other nodes. Dynamic building, mining, and developing of the semantic associations among resources are vital and constitute the objectives of resource evolution. The final objective of building associations between resources is to promote effective learning. Associations between resources can help produce a resource network that is infinitely expandable. The technology used for the automatic building of semantic associations will overcome the limitations of manual manipulation, such as high levels of subjectivity and demands of time and labor.

**Evolutionary mechanisms and technologies in OKCs**

In the current situation, most current OKCs, such as Wikipedia, Hudong Baike, and Cohere (http://cohere.open.ac.uk/), still use the manual control approach to check content quality and achieve the orderly evolution of learning resources (Yang, 2012). Content version management is a commonly used technology. However, this method is time and labor consuming with unsatisfactory efficiency. To reduce manual labor and accelerate resource evolution in OKCs, some studies have been recently conducted to improve the mechanisms of resource evolution in Wikipedia with the use of trust evaluation technology. Adler et al. (2008) proposed a system...
that computes quantitative values of trust of Wikipedia articles. Maniu et al. (2011) studied the signed web of user trust from user interactions. The values of article trust and user trust can provide an indication of reliability. However, only a few studies have been further conducted to control content evolution automatically. In addition, Vrandecic (2009) proposed the use of semantics to check content quality automatically. Although controlling resource evolution in OKCs is a good idea, Vrandecic failed to not propose any feasible technical solution. Therefore, a new and intelligent method is required to control content evolution automatically in OKCs.

Regarding the technologies of association evolution, some studies have been conducted in recent years in addition to the method of manually editing hyperlinks. At present, the learning object (LO) is the main resource form in the e-learning domain. Some scholars studied the association technologies of LOs, which mainly involve relationship metadata design (Ullrich, 2005; Lu & Hsieh, 2009), association representation design (Lv & Du, 2010; Shi et al., 2003), similarity measurement (Zhang et al., 2006), association path search (Li, 2005), and automatic assembly (Farrell et al., 2004). Generally, current research focusing on association technology has two shortcomings, as follows: (1) Representation of the association relationship mainly adopts static metadata description technology without considering the semantic relationship among resources and without standardized methods to describe resource association. (2) Although some researchers have considered the calculation of semantic relationships from the perspective of resource ontology, they are mainly restricted to similarity relationship measurement, ignoring the dynamic nature of semantic relationships (e.g., preorder, successor, and opposition) of other resources. Therefore, we should enhance the relationships between learning resources from the perspective of semantics in OKCs and examine a new technical solution.

Research objectives and questions

In summary, content evolution control and semantic associations of learning resources have become the two major practical issues in the development of OKCs (Yang, 2012). To address the above issues and promote the orderly evolution of learning resources in OKCs, we propose a resource evolution support system (RESS) for open knowledge communities. Two major research questions are as follows:

- How can the direction of content evolution be intelligently controlled in the process of multiuser collaborative content editing?
- How can rich semantic associations be built among learning resources in an automated manner?

Functional architecture

A learning cell system (LCS, http://lcell.bnu.edu.cn) is an ideal platform for a study of this nature because it is specifically developed to support open social learning. LCS provides an open application–programming interface for third-party developers. This platform also supports collaborative authoring and ontology management, which saves time and maximizes efficiency. In this study, we develop the RESS based on LCS. We then provide an overview of LCS and describe the system architecture design of RESS.

Overview of LCS

LCS is an OKC for open learning (Yu et al., 2015). The platform consists of six functional modules: Learning cell (LC), knowledge group (KG), knowledge cloud (KC), learning tool (LT), learning community (LCm), and personal space (PS). Unlike the ad hoc collaboration of Wikipedia, users actively collaborate on a learning resource to achieve a common learning objective.

LC is a resource entity that can be a lesson or a knowledge point. Each LC contains not only content but also learning activities, semantic information, and generative information. The LC can introduce related assistant LTs to support learning. Each KG consists of LCs on related subjects. For instance, a course can be a KG, and each lesson or knowledge point in the course can be an LC. When users access KG, they can find all of the LCs related to the course. The KC aggregates multiple KGs. Different KGs are connected via semantic relationships. In a KC, users can easily find all of the KGs related to their subject. The LT assembles all of the personalized learning gadgets, which can be used by LCs, KGs, PSs, and LCms. An LCm is a collective learning environment in which community
members collaborate and share with one another. For instance, members can publish a notice, initiate a discussion, share resources, and initiate learning activities. In addition to LCs, all users have their own personalized learning environment (PLE). In the PLE, users can post basic personal information, manage (create, collaborate, and subscribe to) LCs and KGs, and select recommended learning resources.

All users can edit and improve the resource content within the framework of collaborative editing (Figure 3), which is crucial for sustaining the evolution of content. To ensure security during evolution, a content version management function was implemented in LCS.

LCS resources are organized using ontology technology. LCS includes three different ontologies, namely, knowledge, user, and context ontology. An ontology management module was developed based on the JENA framework. Unlike traditional ontologies created by experts, LCS adopted an open and collaborative technique to construct discipline ontologies. Any user from a particular disciplinary background was permitted to create the corresponding discipline ontology. In addition to manual construction, some core discipline concepts were extracted automatically from resource contents to enrich the ontology base. External ontology bases (e.g., FOAF (http://www.foaf-project.org/), vCard (http://www.foaf-project.org)) could be quickly imported, and the ontologies generated in LCS could be exported to standardized web ontology language files for sharing with other systems. Ontology refining was implemented to remove outdated and unqualified concepts and properties automatically, as well as to guarantee the quality of ontologies.

**System architecture design**

The functional architecture of RESS in LCS is shown in Figure 4. Technologies, such as Java 2 Platform, Enterprise Edition, ontology, trust evaluation model, rule-based reasoning, association rule mining (ARM), and Flex were adopted. The three main modules are content evolution control, semantic association, and visual presentation of the resource evolutionary path.
Content evolution control

The core function of this module is to take intelligent control of the direction of resource evolution by automatically reviewing new versions of content. Semantic gene abstraction and trust computing underpin evolution control. The semantic gene is the basic information unit representing the meaning conveyed by the resource content. Formally, the semantic gene appears as a set of concepts in domain ontologies with assigned weights and as a set of semantic associations among these concepts. The trust evaluation model is commonly used for assessing the trust level of any entity involved in network interaction. Trust evaluation is modeled on trust relationships in society. By effectively integrating the semantic gene and user trustworthiness, the system could automatically conduct content version auditing, with the process stored in system logs (Figure 5). The statistical function of evolution control could help the system administrator in mastering the overall effect of intelligent content evolution control.

Dynamic semantic association

Dynamic semantic association differs from webpage hyperlinks, which have no semantic association. The module automatically establishes standardized semantic relationships among resources. Association relationships among resources can vary and need to be updated and developed with each change of content. Lu and Hsieh (2009) identified 15 new semantic relationships based on the metadata in SCORM (Sharable Content Object Reference
Model). Using these relationships, 32 semantic relationships were used for building resource associations. The system administrator has rights to manage all of the relationships. Learning resource associations can be automatically established, and the results of resource associations can be analyzed statistically and presented visually. The system provides three ways to build semantic associations among resources using semantic gene computing technology, rule-based reasoning technology, and ARM technology. Figure 6 shows the semantic association network of LCs in one KG.

![Semantic network](image)

The node represents the learning resource.

The line represents the relationship between two resource nodes.

**Figure 6.** Semantic associations among learning resources

**Visual presentation of the evolutionary path**

Flex technology was used to visualize the resource evolutionary path and to present the process of resource evolution more intuitively. Learners could view the history of resource improvement and progress, which may be helpful for learning. The changes of resource content and associations at different times were presented visually (Figure 7).

![Visualization](image)

**Figure 7.** Visual presentation of the resource evolutionary path
Implementation technologies

Corresponding to the aforementioned specific research questions, we propose two solutions by combining technologies of semantics, trust evaluation, rule-based reasoning, and ARM.

Intelligent control of content evolution

The intelligent control method of content evolution is designed based on two assumptions, as follows:

Assumption 1: The evolution of content centers on specific subjects, with strong semantic associations usually linking content before and after changes are made.

The content evolution of resources usually shows a clear direction and is linked to the development of surrounding specific knowledge structures (i.e., the semantic gene). The content introduces a particular subject, which is then updated to gradually enrich and perfect the subject. These “before and after” changes to content usually have strong semantic associations. As such, the newly added content and semantic gene are likely to be semantically similar.

Assumption 2: The behavior of highly trusted users is reliable and inclined to well-meaning content editing.

The trustworthiness of a user derives from a trust evaluation model and is computed by analyzing various interaction data. As a user accumulates desirable behavior, their trustworthiness improves. Once the trustworthiness of a user exceeds a certain value, most operations of that user are deemed reliable and their content editing will be automatically accepted. Figure 8 shows the technical framework for implementing the intelligent control of content evolution based on the semantic gene and the trust evaluation model.

Figure 8. Procedure of intelligent control of content evolution. Note. AT, acceptance threshold.

The core idea is to compute the credibility of content edits by integrating two types of information. The first type is the semantic similarity between the semantic gene of the current learning resource and the feature set of the newly added content. The second type is to compute the trustworthiness of the content editor using the trust evaluation model based on key interactive operational data. If the computed credibility level of content editing exceeds the preset threshold, then the content edit will be automatically accepted; if not, it will be rejected. The creator of the
resource has rights to modify the audit results manually to guarantee the quality of evolution control. The system will automatically accept or reject the content editing and immediately send e-mail notifications to the editor and manager. This feedback mechanism could be helpful for expediting resource evolution.

The semantic gene can be represented through concept aggregation, involving the core concept and relationship between concepts. Weighting may be given based on a description of ontology. The semantic gene can be presented using the following three-element format: \( SG = \langle CS, WS, RS \rangle \), where \( CS \) is the aggregation of the core concept, \( CS = \{C_1, C_2, C_3, \ldots, C_n\} \); \( WS \) is the weighted aggregate of concept item, \( WS = \{W_1, W_2, W_3, \ldots, W_n\} \); \( W_1 \) is the weighting of \( C_1 \); and \( RS \) is the relationship aggregation among core concepts, \( RS = \{R_1, R_2, R_3, \ldots, R_n\} \). Each relationship is presented using a resource description framework (RDF) triple \( \langle Subject, Predicate, Object \rangle \) of domain ontology, where \( R_1 = \langle Concept_1, Relationship, Concept_2 \rangle \). In this study, \( Concept_1 \) and \( Concept_2 \) may not be contained in the CS and could be concepts from other domain ontology bases. \( Relationship \) is the concept relationship abstracted from the domain ontology base.

To abstract a semantic gene from learning resource content, the resource entity should first be structurally represented. The entity can be represented with four parts, as follows: \( Res = \langle Title, Tag, Content, SemanticData \rangle \), where \( Title \) is the title of the resource, \( Tag \) is the tag added to the resource, \( Content \) is the detailed content of the resource, and \( SemanticData \) is the ontology-based semantic description information attached to the resource. \( Title \), \( Tag \), \( Content \), and \( SemanticData \) are the four main sources for abstracting the semantic gene and have different relative weighting for representing the core content of a resource. Weighting aggregation can be presented as \( WT = \{WT_1, WT_2, WT_3, WT_4\} \), where \( WT_1 \) is the weighting of \( SemanticData \), \( WT_2 \) is weighting of \( Title \), \( WT_3 \) is weighting of \( Tag \), and \( WT_4 \) is weighting of \( Content \). After resolving the source and weighting, the characteristic item abstraction technology from the web data mining domain is combined with a series of characteristic words (core concepts) extracted by the domain ontology base. Then, these words are mapped to ontology and stored in CS aggregation. Each characteristic is given a different weighting value using a preset characteristic evaluation function (see Formula (1)). These weighting values are placed in WS aggregation. Finally, using the JENA framework, the semantic relationship of these characteristic words in the domain ontology base is abstracted and placed in RS aggregation.

\[
FE(i) = \log(CF(c, SemanticData) \times WT_1 + CF(c, Title) \times WT_2 + CF(c, Tag) \times WT_3 + CF(c, Content) \times WT_4)
\]  

(1)

In Formula (1), \( CF(c, x) \) represents the frequency of occurrence of concept \( c \) in \( x \), \( x \in \{SemanticData, Title, Tag, Content\} \). Notably, the initial semantic gene comes mainly from the basic concepts extracted from the title of the learning resource. Along with the development of resource content, its semantic gene will become more accurate and rich.

The text features of new content are extracted using the method proposed by Ray and Chandra (2012). The process involves the adoption of a key word set using statistical methods to calculate the weighting of characteristic words via the characteristic evaluation function. Ultimately, a key word set with appropriate weighting is produced. In contrast to the semantic gene of a resource, the text characteristic of newly added content lacks RS. The semantic similarity calculation of two separate texts requires the use of a vector space-based cosine algorithm (Jin, 2009).

In addition to the semantic gene, the trust evaluation model is another core element in the content evolution control technical framework. Many trust evaluation models have been developed for network communication and electronic business (Jones & Leonard, 2008; Li & Wang, 2011; Denko et al., 2011). Considering the different interactions in OKCs, some researchers have studied trust evaluation methods adapted for OKCs (Rowley & Johnson, 2013; Javanmardi et al., 2010; Moturu & Liu, 2009). However, these studies considered the disadvantages, such as inadequate influencing factors, and separated the treatment of trustworthiness for users and resources. Based on this research, we developed a two-way interactive feedback model (Yang et al., 2014). The model has two core components, namely, resource trustworthiness (RT) and user trustworthiness (UT). The model is based on more interaction data, considers the interrelation between RT and UT, and better represents the features of interpersonal trust. In this study, we limit ourselves to a description of the computing method of user trustworthiness. More details can be found in the study of Yang et al. (2014).

\( UT \) is described by four elements: \( UT = \{UT_{res}, UT_{col}, UT_{fri}, UT_{rev}\} \), where \( UT_{res} \) is the trustworthiness component for a user calculated from the resources that he/she created, \( UT_{col} \) is calculated from his/her interaction with other users,
$UT_{ri}$ is based on friendship relationships between this user and other community members, and $UT_{rev}$ is a component calculated from his/her editing history in the community. Equation (2) is used to calculate the user trustworthiness by combining $UT_{res}$, $UT_{col}$, $UT_{fri}$, and $UT_{rev}$. The relative importance of the four components is described by a user weight set: $UW = (UW_1, UW_2, UW_3, UW_4)$ ($\sum UW_i = 1$).

$$UT = UW_1 \times UT_{res} + UW_2 \times UT_{col} + UW_3 \times UT_{fri} + UW_4 \times UT_{rev}$$

(2)

**Semantic association of learning resources**

The abundant semantic association among learning resources can enhance the interconnection among resource entities, improve the frequency of browsing or content editing for each resource, and promote the rapid evolution of a resource. The abundant semantic association can also provide the data foundation from which to develop dynamic aggregation leading to resource groups with larger size and internal logic relationships.

In this study, we propose the learning resource dynamic semantic association technical framework shown in Figure 9. The framework implements dynamic semantic association in association evolution. Semantic annotation is conducted on learning resources using the concepts and attributes in a knowledge ontology base to generate a semantic information space that contains a range of standardized semantic description information.

![Figure 9. Semantic association and aggregation in association evolution. Note. SG, semantic gene; RBR, rule-based reasoning; ARM, association rule mining.](image)

To build semantic associations among resources automatically, we used semantic genes to compute the types of relationships among resources. This method requires the use of semantic dictionaries, such as HowNet (Dong et al., 2007) or WordNet (Miller, 1995). To calculate the similarity relationship based on the semantic gene, we first calculated the similarity between two concepts using semantic dictionaries and domain ontology. Then, we set similarity weighting values in combination with weighting values of concepts in the semantic gene. Finally, we determined whether the similarity relationship between two resources met the similarity threshold. The semantic similarity between two concepts is calculated via the 3D-SIM method proposed by Benwusiyi et al. (Wu & Wu, 2010).

Second, we adopted rule-based reasoning technology to distinguish specific associations based on the existing associations in a range of preset rules, which can be dynamically enriched and changed. JENA is a Java framework for building semantic web applications and provides an inference engine to assist in developing reasoning applications. The built-in inference engine of JENA was used to compile the association rules in this study. First, we compiled various association rules that were saved in an inference rule base. Then, the JENA inference engine extracted the rule from the rule base and linked the rule to the ontology model. Finally, the dominant resource association aggregation was outputted. Before adopting the JENA framework to achieve the dominant resource association based on rule inference, two critical processes were completed, as follows: (1) Data were saved in a JENA-supported ontology model via the RDF triple form. (2) Various association rules were compiled subject to the rule form defined by the JENA inference engine. The inference engine binds these rules and conducts inference using the ontology model to derive a new ontology model.
An invisible association is difficult to detect by human eye. The ARM technology, which is an important and frequently used data mining technology, was introduced to determine these potential associations. The ARM can obtain better results than association rules based on domain ontology (Babashzadeh et al., 2013). In this study, we improved the traditional ARM algorithm by adding semantic gene-based constraints, which can enhance the efficiency and precision of mining association rules. In this study, we improved the a priori algorithm by considering the minimum supporting degree (min_supp), minimum confidence (min_conf), and minimum semantic correlation (min_semrel) required to restrict the creation of association rules to improve the efficiency and accuracy of ARM. We proposed a semantic constraint-based association rule mining algorithm (semantic constraint a priori) (Yang et al., 2013). The minimum semantic correlation refers to the minimum similarity between entities contained in frequent items, where min_semrel can be calculated via the semantic gene of the resource. Through min_semrel, many insignificant candidate items can be filtered to improve the efficiency of the algorithm and, ultimately, create an association rule of higher quality.

**Results of system operation**

The RESS was released and ran from 1 February 2012. Then, we analyzed the evolution process of learning resources in LCS aided by RESS, as well as the effects of content evolution control and semantic association of resources.

**Demonstration of resource evolution**

The LC entitled “Overview of Corporate Universities” was selected as a typical case to depict the evolution process. The reasons for this choice were as follows: (1) Corporate universities are currently one of the “hottest” topics in the field of educational technology. As such, many learners may be engaged in improving the LC. (2) The LCS already contains a number of resources related to corporate universities. Therefore, building resource associations and observing the effect of resource associations would be easy.

The LC was created by user maxyang on 10 February 2012. By 6 March 2012, 26 formal versions had been generated. Version 1 contained basic contents related to corporate universities, including their origin, definition, and features. Altogether, 15 users edited the contents of the LC 68 times until Version 26 was produced. Specific revisions included adding new paragraphs, adjusting character styles, embedding photographs and activities, formatting paragraphs, and changing content. Figure 10 shows the result of the intelligent control of the evolution of the LC. At 16:58 on 29 February 2012, user noteexxx added content related to the subject of love. As this deviated significantly from the semantic gene of the LC, the system automatically rejected this edit and notified user noteexxx and the creator of this resource by e-mail. At 19:52 on 29 February 2012, user 李山 (Li Shan) added photographs of corporate universities. As 李山 was a user with a high trust level, the system automatically accepted his contribution.

By 6 March 2012, Version 26 had evolved significantly from the original version, with richer and more precise content. Some new topics and sections had been supplied, such as “Domestic and Foreign Well-known Corporate Universities” and “Reference Materials.” With the aid of Flex technology, the entire evolutionary process was visually tracked, as shown in Figure 7.

In addition to content evolution, resource associations also kept evolving from the outset. When the LC was created, it was not connected with any other resources. As contents evolved, Version 22 had connected with 8 LCs and 16 users. When the LC evolved to Version 26, it was associated with 12 LCs and 29 users. Figure 11 shows the changes in the numbers of resource associations at different evolutionary stages.

The general process of LC evolution is described as follows: (1) A user creates the LC and invites collaborators to edit contents. (2) More users are attracted to participate in viewing, editing, collecting, subscribing, and commenting on the LC. (3) Different versions of resource contents are generated by different users and checked automatically by the intelligent control program. (4) Content versions are constantly updated to achieve the evolution of resource contents. (5) As the contents evolve, the LC slowly and continually connects with other LCs and users through the dynamic semantic association program.
This revision is rejected automatically.

This revision is accepted automatically.

Figure 10. Screenshot of intelligent control of the evolutionary process

The change process of related resources and users

![Graph showing the change process of related resources and users]

Figure 11. Process of association evolution of the LC

Effects of content evolution control

LCS has a log of resource evolution control, which can record the editorial time, reason, editor, audited result, and audit method. In addition, the audit method can be conducted manually or automatically. In a manual audit, the managers of the LCS conduct the checks. An automatic audit uses intelligent control methods.

We randomly selected the one month log data on content editing with a total of 3,938 edit records to check the effect of resource content evolution. A total of 497 edits required auditing, including 340 automatic reviews by the
intelligent control procedures (70.98%) and 139 instances of manual auditing (29.02%). Figure 12 shows that the intelligent control program reduced the workload of content auditing by approximately seven tenths.

Moreover, we investigated the accuracy of the evolving content via intelligent control. One hundred fifty editing records were selected randomly by the program and stored in an Excel datasheet. A Ph.D. student was invited to log on to LCS to evaluate the accuracy of each record. Results showed that, of the 150 edits, 124 were evaluated as correct, yielding an accuracy rate of 82.67%. Focusing on the incorrect audits, we observed that most of these included multimedia contents, such as videos, animations, and pictures. These edits are difficult for the system to deal with. With the growth of registered LCS users, user groups will become larger and user actions may be more difficult to predict. The accuracy rate of intelligent control should be further analyzed, and the methods should be updated based on the results of studies such as this one.

**Effects of resource association**

The two ways to establish associations between LCs are manual editing and automatic building. The automatic building of resource association applies the methods of dynamic semantic association, including rule-based ratiocination, semantic gene-based calculating, and association rule-based data mining (affair of favorites or subscribes).

We developed a statistical function of resource association to monitor the evolution of the overall resource association. The result showed that the resource total (RT) number of resources in LCS was 3,775 and the association total (AT) of resources in LCS was 3,557. The average association degree (AAD; AT/RT) was 0.91, which means that each LC had an average semantic association of 0.91 with other LCs. This ratio implies that the semantic association between resources is not high.

Further analysis shows that 2,918 LCs have large differences in content and do not have semantic association with each other at all. Current resource associations in LCS are mainly distributed in 857 LCs (approximately 22.7%) that form a partial resource association net. The partial AAD number is 4.15. The distribution of resource association in LCS seems to follow a Pareto principle (80:20 rule), that is, over 80% of the total associations are concentrated in 20% of the resources.

**Table 1. Statistics of resource association**

<table>
<thead>
<tr>
<th></th>
<th>Manual</th>
<th>SG</th>
<th>RBR</th>
<th>ARM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>570</td>
<td>783</td>
<td>2,156</td>
<td>672</td>
</tr>
<tr>
<td>Percentage</td>
<td>13.6%</td>
<td>18.7%</td>
<td>51.6%</td>
<td>16.1%</td>
</tr>
</tbody>
</table>

Table 1 shows the statistical result of resource association by different methods. The percentage of manual association is 13.6%, whereas that of automatic association is 86.4% (semantic gene-based calculating = 18.7%, rule-based ratiocination = 51.6%, association rule-based data mining = 16.1%). The results indicate that the associations built by users are few. Most of the associations are automatically created through dynamic semantic association, which plays an important role in the association evolution of the LCS resource group.
In addition, the accuracy of resource association also requires consideration. Considering the entire dynamic semantic association method, we examined the accuracy of the automatically established resource association. One hundred fifty resource association records were randomly selected by the program and stored in Excel datasheets. Two Ph.D. students were invited to evaluate the accuracy of 75 resource associations each and note their evaluations in the datasheets. A kappa consistency test revealed a result of 0.81 (p < 0.01). The result shows that the precision of automatically established associations is 71.33%. This finding indicates that the semantic association method has a high reliability and that most of the associations automatically established between resources are accurate. However, the monitoring result shows that the semantic gene-based similarity relationship calculation has disadvantages in performance efficiency and time demands. Therefore, the algorithm needs further testing and optimization.

Conclusion

In this study, we develop the RESS, which mainly addresses two issues in OKCs, namely, the intelligent control of content evolution and the semantic association. Integrated application of semantic features of resources based on ontology and a trust evaluation model is feasible for automatically checking content edits in the effort to achieve intelligent control of content evolution. Semantic association is an important part of resource evolution. Through relation calculations based on the semantic gene, rule-based reasoning, and association rule mining with semantic constraint, semantic associations could be dynamically built in association evolution.

This study generally makes two evident academic contributions to the area of OKCs. First, a novel RESS, which may shed light on the improvement of various OKCs, is proposed. Second, the integration of technologies consisting of semantics, trust evaluation, rule-based reasoning, and association rule mining creates an effective and novel method by which to implement the intelligent control of content evolution and the semantic association among resources. This approach can be applied to other OKCs that feature collaborative content editing. The limitation of this work is that no comparison is conducted between the effects of traditional methods and those of the new methods proposed in this study in relation to addressing problems of content evolution control and semantic association.

Moreover, the results have some implications for the applications of OKCs in educational scenarios. First, RESS can be used to promote the generation of course content in the activities of SGC. Second, teachers can build the semantic network of knowledge points of courses assisted by the semantic association technology proposed in this study. Third, when searching knowledge in OKCs, teachers and students should consider the credibility of learning resources and related users.

In the future, we plan to track the process of resource evolution continually in LCS, as well as to accumulate more data to assess the evolution effect. In addition, comparison studies will be conducted to assess the actual effectiveness of the methods proposed in this study.

Acknowledgments

This work was sponsored by a project funded by the National Social Science Foundation of China for the Youth Scholars of Education (Grant Number: CCA130134).

References


Impacts of Pedagogical Agent Gender in an Accessible Learning Environment

Noah L. Schroeder1* and Olusola O. Adesope2
1Wright State University, USA // 2Washington State University, USA // noah.schroeder@wright.edu // olusola.adesope@wsu.edu

*Corresponding author

(Submitted August 25, 2014; Revised March 19, 2015; Accepted April 7, 2015)

ABSTRACT
Advances in information technologies have resulted in the use of pedagogical agents to facilitate learning. Although several studies have been conducted to examine the effects of pedagogical agents on learning, little is known about gender stereotypes of agents and how those stereotypes influence student learning and attitudes. This study investigates if the pedagogical agent’s gender influences cognitive and affective outcomes in learner-attenuated system-paced environments. The findings reveal that the gender of the agent did not produce any statistically significant effects. This indicates that the effects of learning with pedagogical agents may be independent of the agents’ gender. The implications for theory and practice are discussed, along with future research directions.

Keywords
Multimedia learning, Pedagogical agent, Gender, Pacing

Introduction
Learning technologies are changing not only the way we learn, but also how we engage and communicate with one another (Archembault & Crippen, 2009). Without a doubt, the digital media produced by such technologies have significant impacts on modern life (Warschauer & Matuchniak, 2010). While digital media produce new challenges in teaching and learning, technology allows for innovation in the classroom (Shapley, Sheehan, Maloney, & Caramanikas-Walker, 2011). For example, innovative technologies have given us pedagogical agents, which are virtual characters visually present in multimedia environments for the purpose of facilitating learning (Moreno, 2005).

While some studies have shown that pedagogical agents hold potential for facilitating learning (Clarebout, Elen, Johnson, & Shaw, 2002; Schroeder, Adesope, & Barouch Gilbert, 2013), their design and implementation must be thoughtfully guided. For instance, Veletsianos, Miller, and Doering (2009) suggested that pedagogical agent implementation should focus on improving “the social, pedagogical, and technological opportunities provided to learners” (p. 179). If these considerations remain absent yet an agent is implemented in a learning system, it is possible that a split-attention effect could occur (Moreno, 2005). In other words, a pedagogical agent’s presence could be distracting to the learner (Moreno, 2005; van Mulken, André, & Müller, 1998), thereby inhibiting learning. Hence, it appears that implementing a pedagogical agent effectively is not only a matter of obtaining the appropriate software, but also appropriately designing the agent to enhance learning.

While the theoretical implementations of pedagogical agents are appealing, one may wonder what practical use they offer. Researchers have claimed that pedagogical agents can model or demonstrate tasks or skills, help coach students, or even provide scaffolding for learning (Clarebout, Elen, Johnson, & Shaw, 2002). Clearly, when articulated and implemented appropriately pedagogical agents provide a versatility rivaled by few technology tools. Further, as demonstrated by previous research, pedagogical agents can be used in a wide range of instructional domains, including the humanities, mathematics, and science (e.g., Atkinson, 2002; Domagk, 2010; Dunsworth & Atkinson, 2007; Johnson, Ozogul, Moreno, & Reisslein, 2013; Kizilkaya & Askar, 2008; Moreno, Mayer, Spires, & Lester, 2001; Ozogul, Johnson, Atkinson, & Reisslein, 2013). Yet, these experiments represent just the tip of the iceberg. The potential uses of pedagogical agents are what provide the most exciting insights. For instance, as technology evolves, pedagogical agents could help learners participate in virtual science experiments in biology or chemistry classrooms, explore virtual environments or museums in history or government classes, or even demonstrate the customs of different cultures for those learning a second language or planning to travel abroad. It is plausible that as one technological innovation leads to the next, the role the pedagogical agent plays will be limited only by the software’s abilities and by the designer’s imagination.
Yet, one aspect of pedagogical agent implementation that deserves thorough investigation is the effect of the agent's physical appearance (Domagk, 2010; Veletsianos, 2010). Researchers have found that learners stereotype pedagogical agents by their physical appearance and non-verbal cues (Moreno et al., 2002; Veletsianos, 2010). This study investigates one particular facet of this interaction: gender stereotypes and their impact on learning and perceptions. Gender stereotypes are unconscious thought processes that guide expectations of how each gender should look, speak, and behave (Llorente & Morales, 2012; Erdin, 2009). Accordingly, it is easy to see how gender stereotypes could influence an interaction between the agent and the learner. For the purpose of this study, we are interested in how students project stereotypes onto their instructor, or in this case, how gender stereotypes come into play due to the gender of the pedagogical agent.

Another important aspect of pedagogical agent design is the environment itself. More specifically, the pacing of the environment deserves to be more closely examined. Many pedagogical agent studies have utilized learner-paced environments, where the learner can move forward or backward through pre-determined segments of instruction (e.g., Dunsworth & Atkinson, 2007). Alternatively, very few studies have examined system-paced environments, where the learner had no control over the speed or presentation of the information. In other words, the learner is shown a video clip and cannot pause or rewind it.

Anecdotal reflection in the classroom reveals that not all learning systems can be split into this clean dichotomy of system- or learner-paced. For example, consider a non-segmented video that allows learners to pause, fast-forward, and rewind, such as an individual video clip from the Khan Academy (https://www.khanacademy.org/) or a recorded lecture (e.g., Pierce & Fox, 2012). These types of videos are becoming increasingly common in the classroom as faculty members harness the power and accessibility of the Internet. For instance, the notion of a “flipped classroom” is becoming more popular (Tucker, 2012) and faculty members are recording their presentations or lectures and uploading the videos onto the Internet for students to access outside of class (e.g., Crampton, Vanniasinkam, & Ragusa, 2012; Pierce & Fox, 2012; Zappe, Leicht, Messner, Litzinger, & Lee, 2009).

Instructional videos such as those described by Pierce and Fox (2012) and Zappe et al. (2009) do not fall within the theoretical boundaries of learner-pacing as reflected within pedagogical agent research because they are not segmented as in other studies. In other words, while units of instruction may be separate videos, each instructional unit is one contiguous presentation that the student could pause, rewind, or fast-forward. Since the learners had limited control over the presentation, the video also does not fall within the bounds of system-pacing. After consulting the pedagogical agent literature, the authors realized that instructional videos such as the one described have not been characterized as a particular type of pacing. Since evidence suggests there may be some interaction between the pacing of the learning environment and certain instructional design principles (e.g., the modality effect; see Ginns, 2005) the authors feel it is important to clearly delineate the pacing of instructional videos. As such, we feel a video as described could be more properly considered learner-attenuated system-paced (LASP) instruction. Specifically, in LASP instruction the learner has some aspect of control, yet they are still shown a contiguous video clip. To date, we are aware of only one other published pedagogical agent study that utilized this type of pacing (Schroeder & Adesope, 2013). Due to the inherent differences between LASP, learner, and system-paced environments, it is plausible that there may be effects on learner’s cognitive and affective outcomes.

This study investigated the following research questions to explore how a pedagogical agent’s gender might affect cognitive and affective outcomes in pre-service teachers learning with a pedagogical agent in a LASP learning environment.

- How does learning with a male pedagogical agent affect the learner’s free recall, multiple choice, and transfer scores compared to learning with a female agent?
- How does learning with a male pedagogical agent affect the learner’s perception of the agent compared to a female agent?

**Literature review**

Pedagogical agent research has been grounded in theories from a broad range of disciplines. Below, we discuss the most prominent theories that guided this study.
Social agency theory

Social agency theory denotes how social cues from a multimedia learning environment can activate the same communication rules as conversing with another human (Mayer, Sobko, & Mautone, 2003; Veletsianos, Miller, & Doering, 2009). In other words, the learner tries “harder to make sense of what the computer is saying by engaging in deep cognitive processing” (Mayer, et al., 2003, p. 419). Social agency theory has been well supported throughout the literature. For instance, eye-tracking research found that agents were seen as conversational partners (Louwerse, Graesser, McNamara, & Lu, 2008), researchers found that pedagogical agents were seen as social models (Kim, Baylor, & Shen, 2007), and Moreno, Mayer, Spires, and Lester (2001) found that “students learn a computer-based lesson more deeply when it is presented in a social agency environment than when it is presented as a text and graphics source” (p. 209).

Accordingly, it is plausible that the communication between a computer and a learner can be manipulated to foster either instructor-student or student-student interactions. Researchers have found that the visual appearance of pedagogical agents can invoke stereotypes in the learner (Moreno et al., 2002; Veletsianos, 2007). As such, it is also plausible that one could create a specific social interaction by mindfully changing the agent’s features to enhance the agent’s contextual-relevance. These features include the agent’s appearance, behavior, and speech. In the past, Veletsianos (2007) argued that an agent’s relevance could be critical “because it may influence learners’ attention and perceptions and degree of agent relevance, seriousness, and authenticity” (p. 374). This study investigates the effect of learners’ gender stereotypes on their cognitive and affective outcomes by using contextually-relevant agents that varied only by their gender-specific appearance and voice.

Gender stereotypes

Gender stereotypes can create problems in educational contexts (Potvin, Hazari, Tai, & Sadler, 2009) because they influence people’s rationality and how they behave due to certain expectations (Llorente & Morales, 2012). For instance, females are often stereotyped as obedient, shy, submissive, or dependent, while males may be seen as competent, assertive, or aggressive (Llorente & Morales, 2012; Erdin, 2009). Without a doubt, these stereotypes in educational contexts could cause damage to students’ self-esteem or motivation to learn (Erdin, 2009). Accordingly, the first question we must consider is how these stereotypes play out in the classroom.

Classroom interactions can be influenced by gender stereotypes not only from the teacher’s perspective, but also the student’s (Madrid & Hughes, 2010). For instance, teachers may provide male or female students with differing levels of instruction due to their own lack of awareness, expectations, or gender (Davis & Nicaise, 2011). Alternatively, the student may also act on gender stereotypes they project on their teacher. For example, one study found that students felt female teachers provided more student support, while male teachers use an authoritarian teaching style (Madrid & Hughes, 2010). Yet, the literature has not provided consistent, conclusive findings that female or male teachers are thought to be better teachers in regards to learning. Could this be due to students’ perceptions and stereotypes changing as they age? This question is outside the scope of this study, but it is certainly worthy of future research.

Few researchers have examined how gender stereotypes affect learning and perceptions with pedagogical agents. Moreno et al.’s study (2002) showed that the agents were stereotyped by the learners and that male agents were more effective at fostering learning outcomes than female agents. Further work by Kim, Baylor, and Shen (2007) found that a male pedagogical agent was rated higher by learners on affective measures. Yet, not all research in the area has produced consistent results. For instance, Baylor and Kim (2004) found that female agents were able to increase a learner’s self-efficacy more than male agents, and the learning outcome results from a number of experiments has not provided clear conclusions as to one gender being significantly better for learning than the other (Baylor & Kim, 2004; Kim, Baylor, & Shen, 2007). In sum, the limited research surrounding pedagogical agents and gender stereotypes shows that while affective measures seem to be influenced by gender stereotypes, cognitive measures do not. This study sought to expand the literature surrounding pedagogical agent gender by investigating whether contextually-relevant female or male pedagogical agents provide any affective or cognitive advantage in a LASP learning environment.
Methods

Participants and design

The participants in this study were 77 pre-service teachers at a large, public university in the Pacific Northwestern United States. Participants were randomly assigned to work with either a female agent (Figure 1) or male agent (Figure 2) and participated for course credit. Forty participants were in the male agent group, and 37 participants were in the female agent group.

The participants’ average age was 20.75 (SD = 1.60) years old. On average the participants had completed 2.5 years of post-secondary education, and 74 percent of the participants were female. About 86 percent of the sample was Caucasian, 8 percent reported multiple ethnicities, and 4 percent was Hispanic while two percent of the participants chose not to report their ethnicity.

While the participants sometimes use multimedia in their own teaching (M = 3.35, SD = .96, where 1 is “never” and 5 is “almost always”), 71% of the participants had never received formal instruction about multimedia learning theory. The pre-test reiterated this point as the average pre-test score was 4% (M = .55, SD = .82, points possible =
Taken together, the participants’ self-reported experience with formal multimedia learning theory instruction and their pre-test scores showed that they had very little knowledge of multimedia learning theory before the intervention. This was done to ensure that we do not have a ceiling effect relative to the materials used.

Procedure

The study was conducted in a computer laboratory that has 30 identical Dell computers. Each participant used a set of headphones that were attached to the computer before they arrived. The screen resolution was set to the university’s default setting of 1280x1024. When the participants arrived, they were given a piece of paper that contained their user ID. The computer program we developed used the user ID to route each participant into the appropriate experimental condition. The participants were introduced to the experiment and given an opportunity to opt-out if they chose to. Participants then completed the experiment using their assigned computer station.

Materials

Pedagogical agent and learning environment design

When examining a multimedia learning environment, one easily overlooked design feature is the background, or in this case, the virtual environment in which the agent is situated. In many studies, the background environment is a blank screen (e.g., Atkinson, 2002; Domagk, 2010). Yet, Veletsianos (2007) made a compelling argument for the contextual-relevance of the agent. Following this logic, we have expanded previous research by utilizing a background that appears to be a virtual classroom. In order to minimize potential confounding variables, the virtual classroom did not contain any other agents except the peer agent who was providing instruction. However, the classroom background contained desks, chairs, and windows (Figure 1). The background environment was identical between the two conditions.

The learning environment was created using Xtranormal. Xtranormal was an Internet-based program that allowed for the incorporation of pedagogical agents into virtual learning environments. The program was designed to be extremely user-friendly, as there was no required coding or special software. Rather, one utilized mostly drag and drop methods to create their instructional materials. Videos created with Xtranormal afford learners some aspect of control, similar to videos on other popular educational websites (e.g., Khan Academy, https://www.khanacademy.org/), thus placing them within the context of LASP instruction. Since the time of this study Xtranormal has closed, and the software is now under new ownership (http://www.nawmal.com).

Finally, as has been common practice in pedagogical agent research, the agent in this study provided the instructional material to the learner. Since this study introduced a new type of pacing to pedagogical agent research, as well as expanded the research around peer agents (discussed in the next section) and added a contextually-relevant background scene we did not deviate from the traditional role of the agent as the primary source of information. While delivering instruction, the agent made five gestures. While the gestures did not reference any on-screen learning materials, they were added in the hope of aiding understanding (Hostetter, 2011). Furthermore, the gestures were designed to aid in the agent’s deictic believability, or its ability to move in relation to objects in the virtual space (Lester, Voerman, Towns, & Callaway, 1999), which presumably may facilitate its ability to be perceived as human-like.

Peer agents

Researchers have primarily used instructor-type agents rather than peer agents in pedagogical agent studies (Clarebout et al., 2002). However, it is hard to argue with the compelling evidence for peer interaction within classroom environments (Kim & Baylor, 2006, p. 569). Kim and Baylor (2006) argued that peer agents “may serve as a social model for enhanced motivation and learning in computer-based environments” (p. 580), and be able to foster a learner’s efficacy beliefs. These notions suggest that peer agents may foster both cognitive and affective outcomes more effectively than instructor agents. As such, we question if peer agents may foster an increased sense of social agency than an instructor agent. While researchers have recently utilized peer agents to foster learners’
attraction to an unpopular knowledge domain (Kim & Wei, 2011), peer agents are still not often utilized throughout
the literature. Accordingly, this study adds to the growing literature surrounding pedagogical agents by utilizing
contextually-relevant, peer agents.

**Demographic questionnaire**

Demographic information was obtained with a questionnaire consisting of eight questions that addressed the
student’s age, gender, ethnicity, and previous experiences with multimedia and multimedia learning theory.

**Pre-test**

The pre-test consisted of three free response questions. One question addressed cognitive load theory, another
addressed the modality principle, and the final question addressed the split-attention principle. The pre-test was
worth a maximum of 14 points, with one point given for each correct answer. The participants earned one point for
correctly identifying or describing each of the following ideas in their response to the first question: working
memory, long-term memory, schema, germane cognitive load, intrinsic cognitive load, and extraneous cognitive
load. Thus, the first question had 12 possible points. Participants earned one point for correctly describing the split-
attention principle in the second question, and one point for correctly describing the modality principle in the third
question. The goal of the pre-test was to examine the learner’s prior knowledge while minimizing any error due to
guessing.

**Post-test**

The cognitive aspects of the post-test consisted of three different types of questions. First, the participants answered a
free recall question that asked them to write down everything they could remember from the instructional video.
There were a maximum of 18 points awarded for the free recall question. Points were allotted for correctly
identifying (1 point) and describing (1 point) the following ideas: germane cognitive load, schema, extraneous
cognitive load, intrinsic cognitive load, cognitive load theory, long-term memory, working memory, the modality
principle, and the split-attention principle.

The participants then answered 30 multiple-choice questions in which each correct answer was worth one point. The
questions in this section required the participants to recall specific information as well as apply their knowledge to a
hypothetical situation. The scale’s internal consistency reliability was found to be $\alpha = .70$.

The final question was a free response transfer question. The participants were asked to design a lesson plan utilizing
cognitive load theory, the split-attention principle, and the modality principle. The free response transfer question had
a maximum score of 18, with points given for each correct reference to theory when describing their lesson plan. The
grading for this question was identical to that of the free recall question.

To measure the participants’ perceptions of the agent, they completed the Agent Persona Instrument (Ryu & Baylor,
2005). The instrument consisted of 10 items which addressed how well the agent facilitated learning ($\alpha = .94$), five
items which addressed how credible the agent was ($\alpha = .92$), five items which addressed how human-like the agent
was ($\alpha = .87$), and five items which addressed how engaging the agent was ($\alpha = .86$) (Ryu & Baylor, 2005). The
participants responded using a 5-point Likert scale, where 1 is “Strongly Disagree” and 5 is “Strongly Agree”. For
this study, the scale’s internal consistency reliability was found to be $\alpha = .94$ for facilitated learning, $\alpha = .87$ for
credibility, $\alpha = .88$ for human-like, and $\alpha = .89$ for engaging.

**Results and discussion**

The purpose of this study was to investigate if changing a contextually-relevant, peer pedagogical agent’s gender in a
LASP learning environment would influence either cognitive or affective outcomes. This section describes the
findings of the study in relation to previous research.
Research question one: How does learning with a male pedagogical agent affect the learner’s free recall, multiple choice, and transfer scores compared to learning with a female agent?

The data (Table 1) were first examined graphically in order to check for normality (Tabachnick & Fidell, 2013). Multivariate analysis of variance (MANOVA) was conducted using the cognitive tests as the dependent variables and the agent’s gender as the independent variable. Further analyses indicated that the assumption of homogeneous covariance matrices was met (Box’s $M = 3.67, p > .05$), as was the assumption of the equality of error variance for each cognitive measure (Levene’s tests $p > .05$). The assumptions of MANOVA having been met, the analysis revealed no significant differences between groups on cognitive outcomes (Wilks’ $\lambda = .94$, $F(3, 73) = 1.60, p > .05$).

<table>
<thead>
<tr>
<th>Cognitive outcome results</th>
<th>Male ($n = 40$)</th>
<th>Female ($n = 37$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>0.63</td>
<td>0.46</td>
</tr>
<tr>
<td>Free recall</td>
<td>6.68</td>
<td>4.69</td>
</tr>
<tr>
<td>Multiple choice</td>
<td>17.08</td>
<td>16.03</td>
</tr>
<tr>
<td>Transfer</td>
<td>3.20</td>
<td>3.32</td>
</tr>
</tbody>
</table>

Note. Maximum possible scores: Pre-test: 14; Free Recall: 18; Multiple Choice: 30; Transfer: 18.

These findings support some of the literature around pedagogical agent gender. For instance, Baylor and Kim (2004) found no significant differences in learning outcomes between those who worked with female or male pedagogical agents, as did one experiment in Kim, Baylor, and Shen’s (2007) work. Yet, other experiments have found that male pedagogical agents were more effective for learning (Kim, Baylor, & Shen, 2007; Moreno et al., 2002). Accordingly, it is plausible that in some situations learners may show gender stereotypes with pedagogical agents. However, more work is needed to determine the reasons why these stereotypes may occur.

These findings provide meaningful implications for social agency theory. While social agency theory predicts that social cues in multimedia messages can engage the social conversation schema, and thus improve learning (Louwerse, Graesser, Lu, & Mitchell, 2005; Mayer, Sobko, & Mautone, 2003; Moreno et al., 2001), this study has shown that both male and female peer agents can provide statistically equivalent social cues in a LASP learning environment. As such, it appears as though gender stereotypes did not influence the way the social cues were interpreted by the learners. However, since this study was examining only the gender-specific appearance and voice of the agent, both conditions utilized the same gestures and movements throughout the instructional video to improve the internal validity of the study. Future research should investigate if gender-specific movements or gestures impact students’ learning.

While we found that there were not significant differences in learning outcomes between those who worked with the male or female agents, these results have important practical implications. Earlier, we argued that pedagogical agent design must be carefully guided to suite their purpose. Findings from this study show that if the purpose is to facilitate learning in a LASP environment, the gender of the contextually-relevant, peer pedagogical agent may have no significant difference on learning outcomes. However, we caution that this finding may not be replicated in other knowledge domains or for agents that do not appear as peers. For instance, Schroeder et al. (2013) found that pedagogical agents were more effective at facilitating learning in scientific knowledge domains than within the humanities, and Veletsianos (2010) found that agents were stereotyped differently depending upon if they were providing instruction on music or science. Accordingly, future research should explore if an agent’s gender influences learning or affective outcomes in other knowledge domains.

Research question two: How does learning with a male pedagogical agent affect the learner’s perception of the agent compared to a female agent?

We began investigating this research question by first examining the data distribution for each affective scale to examine normality (Tabachnick & Fidell, 2013). The data (Table 2) were found to be normally distributed, so we proceeded with the MANOVA using the agent’s gender as the independent variable and the affective measures as the dependent variables.
Table 2. Affective outcome results

<table>
<thead>
<tr>
<th></th>
<th>Male (n = 40)</th>
<th></th>
<th>Female (n = 37)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Facilitated learning</td>
<td>26.18</td>
<td>8.38</td>
<td>26.62</td>
<td>9.61</td>
</tr>
<tr>
<td>Credible</td>
<td>16.28</td>
<td>3.97</td>
<td>15.73</td>
<td>4.79</td>
</tr>
<tr>
<td>Human-like</td>
<td>10.43</td>
<td>4.19</td>
<td>10.57</td>
<td>4.57</td>
</tr>
<tr>
<td>Engaging</td>
<td>11.15</td>
<td>4.46</td>
<td>11.03</td>
<td>4.37</td>
</tr>
</tbody>
</table>

Note. Maximum possible scores: Facilitated Learning: 50; Credible: 25; Human-Like: 25; Engaging: 25.

Initial analyses indicated that all assumptions for MANOVA were met (Box’s $M = 15.58$, $p > .05$; Levene’s tests $p > .05$). The analysis revealed no significant differences between groups (Wilks’ $\lambda = .98$, $F(4, 72) = .34$, $p > .05$). In other words, the gender of the pedagogical agent did not have a significant impact on how it was perceived by the learners.

The results do not align with the literature surrounding how students’ stereotype those providing instruction. For example, since men may be stereotyped as being authoritative teachers (Madrid & Hughes, 2010), then it is plausible that male agents would receive higher ratings of credibility. Further, since women may be seen as being more supportive teachers (Madrid & Hughes, 2010), one may expect them to receive higher ratings on the facilitated learning scale. Kim, Baylor, and Shen (2007) found that male pedagogical agents were found to be more interesting than female agents, and they were also perceived more positively. Yet, our findings show no significant differences between how agents of different genders were perceived in a LASP environment. Could this be due to the fact that the agents appeared as peers rather than instructors?

As mentioned previously, Veletsianos (2010) found that agents were stereotyped differently depending upon the domain of the learning materials. In this case, humanities materials in relation to learning theory were used. We question if stereotypes would have been reflected in the outcome measures if the materials fell within the domains of science or mathematics. Future research can explore this question.

Our findings also have implications for social agency theory. Based on the literature around gender stereotypes, some may posit that agent gender may influence students’ perception of social cues. Yet, in a LASP environment with contextually-relevant peer pedagogical agents this was not the case. In other words, either gender of pedagogical agent was able to provide the same level of social cues in their multimedia message to the learner in this study.

Limitations

As with any research study, this one is not without its own set of limitations. The most prominent limitation of the study was the voice that provided the narration in both conditions. The voice was created using a text-to-speech generator provided by the Xtranormal program. Unfortunately, the voice was computer-generated, and we could not control the speed of the voice or its inflection. In fact, a few participants stated that the voice was distracting, even though they were not prompted to provide their opinion on the subject. We question if the results of this study would have been the same if recorded, human voices of standard accent had been used rather than using the text-to-speech feature. Future research may explore this.

Veletsianos (2010) noted that a pedagogical agent’s appearance could influence how the agent is perceived. As such, another limitation of this study was the pedagogical agents’ appearance. The agents that were chosen wore gender-specific clothing. However, the male agent was dressed more formally than the female agent. Could these differences in the formality of clothing have caused stereotypes to occur, or rather, not to occur? The results of our study reify Veletsianos’s (2010) call for research surrounding the appearance of pedagogical agents and its effect on cognitive and affective outcomes.

In this study we were not interested in how male and female students stereotyped agents of the same or opposite gender specifically. While the similarity-attraction hypothesis suggests this may be a relevant factor to consider (Moreno & Flowerday, 2006), Moreno and Flowerday’s (2006) results did not support this claim. Future research can investigate if an individual’s gender influences their perception of the agent. In addition, our measurement techniques...
were similar to those of other pedagogical agent researchers who investigated gender stereotypes (Kim, Baylor, & Shen, 2007; Veletsianos, 2010). However, the instruments did not directly measure stereotypic views, but rather tried to infer them from perceptive and learning measures. While we chose our methodology to be consistent with literature in the area, future research should utilize stereotype specific measurement techniques using instruments with validity evidence to support their use.

Conclusion

The results show that male and female peer pedagogical agents were not perceived differently by the learners, and that neither was able to more effectively facilitate learning outcomes. One must then question why, if agents are engaged as conversational partners (Louwerse et al., 2008), did gender stereotypes not affect learning or perceptions in a LASP learning environment? We hypothesize that the nature of the LASP system itself may be the cause. Learner-paced systems have built in pauses where the learner must hit a button to continue, while LASP environments do not have these features. Could these mandatory pauses in the learning material allow the learners to apply these stereotypes in learner-paced environments? In other words, it is plausible that in a LASP environment an engaged learner will utilize the full extent of their working memory to learn the information being presented throughout the duration of the instructional time. This is due to the information being presented as one contiguous flow of information, unless paused by the learner. Alternatively, learner-paced environments force the learner to pause at pre-determined intervals. During these intervals, it is likely that the learner may reflect upon the information that was just presented, thereby allowing additional working memory capacity to shift towards unconscious thought processes such as gender stereotypes. However, at this point there is no empirical evidence to support this notion. Future research can explore if the pacing of the environment and the application of gender stereotypes are related. In addition, future research should examine what learners actually do when presented with a LASP environment. Do they pause the video? Do they rewind the video? If they do pause the video, how long are the pauses, how do the frequency and duration of the pauses correlate to learning outcomes, and how do they impact working memory capacity and stereotype activation?

Another plausible explanation of why we did not see a stereotype effect comes from examining the data for the affective results (Table 2). Examination of the data indicates that the average rating for the learner’s engagement with the agent was less than 50% of the possible score. Could this lack of engagement with the agent have caused the lack of gender stereotypes? To what extent did the agent’s voice effect the learner’s engagement? Replication of this study using recorded human-voices would be beneficial to examine this limitation.

In sum, this study has shown that pedagogical agent research is still in its infancy with many relevant questions to be addressed. The cognitive and affective responses to the pacing of the system, the agent’s appearance, the role of the agent in the learning environment, and the agent’s ability to interact with the learner are just a few areas of research which are yet to be thoroughly explored. We encourage researchers to continue to explore the varied uses of pedagogical agents in an effort to better understand the best practices and implementations of pedagogical agents. However, we echo Clark and Choi’s (2005) call for research that examines the cost-effectiveness of the technology, and encourage researchers to pursue agent technologies that are easily accessible to many instructors.

References


Value-Added Results for Public Virtual Schools in California

Richard Ford1 and Kerry Rice2*

1Department of Mathematics and Statistics, California State University, Chico, CA, USA // 2Department of Educational Technology, Boise State University, Boise, ID, USA // RFord@csuchico.edu // krice@boisestate.edu

*Corresponding author

(Submitted September 4, 2014; Revised January 19, 2015; Accepted February 16, 2015)

ABSTRACT

The objective of this paper is to present value-added calculation methods that were applied to determine whether online schools performed at the same or different levels relative to standardized testing. This study includes information on how we approached our value added model development and the results for 32 online public high schools in California. Student level California Standards Test results in English Language Arts and Mathematics for over 5,000 online students were analyzed. Mean value added metrics for each school were calculated for 8 courses held during the 2010-2011 academic year. We found that schools of distinction existed in 7 of the 8 course categories.

Keywords

Online learning, Value-added, Virtual schools, K-12, English, Mathematics

Introduction

In 2000 approximately 45,000 K-12 students nationwide engaged in some type of formal online learning course or activity. By 2010 that number had grown to over 4 million (Staker, 2011). The accelerating growth in public online coursework at the K-12 levels elevates the importance of research into the efficiency and effectiveness of online education. Nationally we have a critical and pressing need to expand our knowledge base to facilitate the identification of what works best in online learning environments (Means et al., 2010).

The purpose of this study was to determine whether the virtual schools operating in California in the 2010-2011 academic year produced equivalent or different value added results on standardized tests in English language arts and mathematics. To be precise, we looked at eight specific courses, four in math and four in English language arts, offered at the 32 identified virtual schools whose student level test score data for successive years was provided by the California Department of Education (DOE). (The research was conducted independently by the authors and was not supported nor endorsed by the California DOE.) For this study, virtual schools were defined as those schools in which instruction was delivered entirely or primarily through online methods. The California DOE provides a service that identifies schools that deliver a minimum of 30% of content online. This threshold was too low for our purposes so further identification of schools for this study was accomplished through a comprehensive review of all listed charter school websites for information on their primary delivery method. In particular, we looked for schools whose names reflected some online or electronic component and schools that specifically designated themselves as online in their program descriptions. Given the challenges in defining and categorizing online schools, one limitation of the study is that the schools subsequently included in this report likely do not represent a complete sample of all online schools in California.

Method

The research objective was to identify, in each of the eight separate courses, schools that produced statistically superior value added metrics. The initial data pool consisted of all students who took a math or English language arts California Standards Test (CST) in the spring, 2011 at any one of the 32 identified public online schools. This initial pool was back-mapped by the California Department of Education to retrieve corresponding CST test scores for 2010, regardless of which public school generated the pretest score. Thus any student from the initial pool who was also tested anywhere in California in 2010 would remain in the pool. Those students for whom no pretest score could be retrieved were then eliminated from the pool. In English language arts approximately 82% of the initial pool was retained for the study. In mathematics approximately 77% of the initial pool was retained for the study. Each student
Overview of value added methods

The use of value added methods (VAMs) by schools, districts, and states, now dates back over twenty years. The Tennessee Value Added Assessment Model (TV AAS), a layered mixed effects model, developed by William Sanders and Robert McLean of the University of Tennessee, has been in use since 1991 (Sanders & Horn, 1994). Simply stated, value added methods are a way to measure changes in student performance over time. They continue to be strongly encouraged nationally and are even required for states to be competitive for Race To The Top funding (Corcoran, 2010). As a result there has been tremendous growth in the research base providing analysis of the benefits as well as drawbacks of these new methods. In general these methods are very complex and highly technical and there are concerns that they may be used inappropriately (Condie, Lefgren, & Sims, 2014). In large-scale studies, value added methods “have proved valuable for looking at a range of factors affecting achievement and measuring the effects of programs or interventions” (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012, p. 8). However, they may not be an accurate measure of teacher effectiveness given the wide variety of factors that can affect individual student performance. Indeed, even when applying the similar value-added techniques to the same data sets, different researchers can sometimes generate different results (Briggs & Dominiugue, 2011). One example is a study conducted in 2004 in which a team of researchers led by Carmen Tekwe compared four similar, but different value-added approaches; hierarchical linear mixed models (HLM) with and without student covariates, layered mixed effects models (LMEM), and simple fixed effects models (SFEM). The team claimed to show that the results of the LMEM and SFEM models were different, but highly correlated and concluded the much simpler SFEM model was more desirable (Tekwe, Carter, Ma, Algina, Lucas, Roth, Ariet, Fisher, & Resnick, 2004). This claim was reviewed and disputed by other researchers who stated that the Tekwe et al. (2004) study “relied on a narrow data structure, which may have seriously limited its conclusions” (Doran & Fleischman, 2005, p. 85). The fact is that education data is influenced by an endless variety of factors and will always remain noisy, particularly at the teacher level, no matter how sophisticated and complex the method is. Still the search for ever better models that are fair, comprehensible, and provide reproducible results should certainly continue. As such, one of our long-term goals is to determine if the results from appropriately applied value added calculations could be relevant and reliable at a program level. By identifying those programs or schools that produce exceptional results would it then be possible to tease out the reasons why they were successful?

One of the simplest models we found was used by the United Kingdom to calculate value-added for their schools between 2000 and 2004. This UK method did not take into consideration the wide variety of student, school and other confounding characteristics, which might influence performance. Instead they assumed that, on average, those characteristics are randomly distributed in each of two distinct school classifications, “mainstream” and “special” schools. Rather than using linear regression, they establish a “natural median line” which consists of the set of all points, \((x, y)\) where \(x\) is a particular pretest score range and \(y\) is the median of all posttests scores from students with \(x\) in the pretest range (DfES Analytical Services, 2004). The value added score for any student is the difference between their posttest score and the corresponding median score for their specific pretest value. The value added for a school is then calculated as the mean of the value added scores for all the students in the school, plus 1,000. Thus a school with a score of 995 is below average and one with a score of 1006 is above average. This system was very attractive to policy makers due to its simplicity. It is quite likely that use at the teacher level would not be particularly reliable due to the wide variety of student characteristics that do not distribute evenly at the classroom level. Since
our goal was not to evaluate performance at the teacher level, but rather at the program or school level, we concluded that this level of analysis was similar to what we might need. We note that this model has since been replaced in the UK with a more sophisticated and complex model (Evans, 2008).

**Complex model issues**

The objective of ever more complex value-added models is to control for variables that contribute to student advancement that are unrelated to the teacher or school education inputs. The models we studied (Amrein-Beardsley & Collins, 2012; Atteberry, 2012; DfES Analytical Services, 2004; Evans, 2008; Goe, 2008; Isenberg & Hock, 2011; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Raudenbush, 2004; Sanders & Horn, 1994; Tekwe, et al., 2004; Value-Added Research Center, 2015; Wright, White, Sanders, & Rivers, 2010) typically employed some type of multiple linear regression to accomplish this control. One way to explain how the control is accomplished is in terms of expected posttest outcomes. Suppose in a large population of students it turns out that students with some specific characteristic, say left-handedness, produce generally greater growth from year to year in a particular subject. We recognize that teachers and schools have nothing to do with whether or not a student is left-handed and therefore wish to make our model fairer by controlling for that variable. In addition, we want our expected posttest prediction to be as accurate as possible, and knowing whether or not a student is left-handed would be information that should help. Linear regression that includes this factor essentially improves the prediction of posttest score by including the average effect differential between right and left-handedness. The end result, simply stated, is that the model will produce an *expected* score for a left-handed student that is appropriately more than one who is right-handed with all other factors equal. When value-added is defined as the residual, i.e. the difference between the *expected* and actual posttest scores, we see that the left-handed student would be assigned lower growth than if the factor were not included in the model. In this way when value-added scores are compiled for schools or teachers that have disproportionate numbers of right or left handed students, they will not be rewarded nor penalized for something unrelated to the educational input provided.

The example above helps to illuminate important education issues related to complex value-added models. The model by itself makes no attempt to explain why our lefties perform better. In fact, by including the factor in the model, motivation by schools, teachers, and administrators to study causal factors is reduced because they are not held accountable for that characteristic. It could be as simple as the fact that our student desks at which the tests are taken are all designed for lefties. We should want to study and mitigate the right-left discrepancies, but instead, accounting for it in our model removes the incentive because the lower relative growth for the right-handed group doesn’t lower the value-added calculation for the teacher or school.

Controlling for some factors may also encourage inappropriate adjustments. Suppose we discover that students with tattoos tend to do more poorly on average than the general population. We decide to control for that variable resulting in adjusting value-added scores slightly higher for those students with tattoos. It’s certainly a bit silly, but perhaps a serious administrator decides that it will help his school’s overall value-added score by asking all his students to get tattoos. The example is far-fetched, but currently many models control for free and reduced-price lunch. This means that if the school could qualify more of its existing students, value-added scores would rise slightly only due to the mathematical calculation adjustments.

These issues surrounding ever more complex value-added models need to be understood and discussed by educators and political leaders. Controlling for variables unrelated to education input can and likely will result in unintended consequences. In our choice of a simpler method, we are making the case that we do not want nor were we able, given the data set, to control for extraneous factors – we give an honest depiction of relative performance given the student population and data set available to us.

**Value added model development**

In the end, the struggle to determine the best VAM to apply in this study was essentially decided for us. Due to the fact that only pre and posttest data were provided whereas demographic, individual student characteristics and other possibly confounding data were not provided, our choices were limited. As stated earlier, the pretest data was generated from California Standards Testing (CST) exams given in the spring of 2010. The posttest data was
generated from CST exams given in the spring of 2011. Our original plan was to follow an established procedure using normal curve equivalents (NCEs). This is a process used in early SAS EVAAS analysis where value added was based, essentially, on the change in normal curve z-scores from year to year, using the reference distribution for the full test-taking population in the state (Wright, White, Sanders, & Rivers, 2010). We learned by examining our data that students who all took the same posttest took differing pretests. For example, from the group whose posttest consisted of Algebra 1, students took any one of four different pretests; 7th grade math, 8th-9th general math, Algebra 1 (repeated) or even Geometry. In this report we define a “cohort” to be any collection of students in the study whose pretest-posttest pair is the same. Thus all those students who took the Geometry exam in 2010 and then took the Algebra 1 exam in 2011 form a single cohort. With this definition we see that the Algebra 1 posttest group includes four significant cohorts. Cohorts consisting of fewer than 36 students were excluded from the analysis due to the low correlation coefficients in the linear regression. Appendix A contains technical data for the cohorts included in the study.

California testing policy required that students enrolled in and attending a specific math or English course in academic year 2010-2011 must take the associated CST exam in the spring of 2011. Therefore the exam taken by a student informed us of exactly which course the student was enrolled in. For example, all students who took the Algebra 1 exam in 2011 were also enrolled in the Algebra 1 course at their virtual school during the academic year 2010-2011. Our objective was to establish a value-added score for each course at each school. This meant we needed to develop some way to pool value-added assessments from the various cohorts taking the same posttest. Excluded cohorts consisted of pretest-posttest pairs that represented unusual course sequencing. For example, one excluded cohort consisted of students whose tests indicated they took Algebra 2 and subsequently took Algebra 1.

It took some time to understand the implications of multiple cohorts taking a single posttest, but after some analysis we realized that the typical SAS EVAAS approach using NCEs would not provide a true picture of value-added. The normal process of equating value-added with the z-score changes based on the respective reference distributions simply doesn’t work when you have multiple cohorts. The NCE approach is essentially equivalent to redefining student test score values as the corresponding z-scores earned on each test relative to the distributions for the full populations taking those exams. One reason this method breaks down in our situation is because the cohorts are not random distributions of the pretest population. For example, one cohort consists of students who took the Algebra 1 exam in both years. We would normally expect a student to repeat Algebra 1 only if they performed below expectations on the pretest. So the mean of the pretest scores for this restricted cohort will certainly be much lower than the reference distribution mean for the full population whose pretest was Algebra 1. Similarly, the English language arts cohort of those students taking the 9th grade ELA exam as their posttest and the 7th grade ELA exam as their pretest would generally consist of those students who did exceptionally well on the pretest and skipped 8th grade. For this cohort we would expect a much higher mean on the pretest than the reference distribution. These kinds of variations in cohort pretest averages unacceptably distort the meaning of value added based on NCEs using reference distributions for the full test taking populations.

Based on the above considerations, we selected a standard linear regression method encouraged by the Value Added Research Center (Value-Added Research Center, 2015). We first established expected posttest scores based on linear regression of the known data in each individual cohort. Residuals then formed the value added score for each student. This method is similar to that used early on by the United Kingdom in establishing value-added measures for their schools. Given our research goals and given the data set we had access to, we were confident the method would produce meaningful distinctions between our identified online schools that could also be digested and duplicated by a wide audience.

The model used to establish expected posttest scores was the following:

\[ \hat{y}_{ijk} = \alpha_{jk} + \beta_{jk}x_{jk} + \epsilon_{jk} \]

Here \( j,k \) represents the pretest-posttest pair and \( i \) indicates the student. The \( \alpha \) and \( \beta \) are regression coefficients, \( \hat{y}_{ijk} \) is the expected posttest score, and \( \epsilon \) is normally distributed random error term with mean 0 and constant variance. Value added is interpreted as the residual or difference between the expected posttest score, \( \hat{y}_{ijk} \) and the actual scaled posttest score earned, \( y_{ijk} \). Since the value added is the residual of the regression, the mean of the value added scores
will necessarily be 0 in each cohort. To combine the value added scores for all cohorts with the same posttest we first convert all the residuals to z-scores based on the distributions of the residuals within the appropriate cohort. This may amplify or dampen the value-added within specific cohorts, but is consistent with the idea that fluctuations in variances between cohorts are primarily artifacts of the varying scaled score magnitudes. We then take the z-score equivalent of the residual as the value added for the particular student. The pooled value-added for the cohorts for a specific posttest were then sorted by school. The mean of these scores represents the value-added for the associated course at the particular school. Mathematically, we calculated that value added as follows:

First, we established value added, $z_{ijk}$, for student $i$ for cohort $j,k$: (pretest $j$, posttest $k$):

$$z_{ijk} = \left( y_{ijk} - \hat{y}_{ijk} \right) / \sigma_{jk}$$

where $\sigma_{jk}$ is the standard deviation of the set of residuals, $\left\{ y_{ijk} - \hat{y}_{ijk} \right\}$ over the $(j, k)$ cohort. Then for each school, the value-added in a specific course associated with posttest $k$ is the mean of the values-added for all students in the study who took the $k^{th}$ posttest at the school.

Standard one-sided 95% confidence intervals were then established for each (course, school) pair. We identified “distinguished schools” for a specific course as any school whose value-added confidence interval was entirely positive. (Note the overall course means will also be zero since each cohort mean is zero.) Our interpretation is that any school that is designated distinguished for a course is above average with 95% certainty.

Data and results

The eight courses studied included the following: English language arts for grades 8, 9, 10 and 11, General Mathematics, Algebra 1, Geometry, and Algebra 2. The distinguished schools, courses, and relevant data are presented below:

English language arts

iHigh Virtual Academy in San Diego had an outstanding 10th grade class in AY 2010-2011. Their adjusted value added was .765 standard deviations above the mean giving 95% confidence that they performed at least .355 standard deviations above the mean on average. They had 18 students, which most likely represents a single class. CA Virtual Academy at Los Angeles performed above average with 95% confidence in both 9th and 10th grades. Their student count is quite high, 262 and 239 respectively. The same was true for CA Virtual Academy at San Diego who performed above average in both 8th and 9th grades.

Distinguished schools performing above average with 95% confidence

ELA 8
CA Virtual Academy, San Diego

ELA 9
CA Virtual Academy, Kings
CA Virtual Academy, LA
CA Virtual Academy, San Diego

ELA 10
Capistrano Connections Academy
CA Virtual Academy, Kings
CA Virtual Academy, Los Angeles
CA Virtual Academy, San Joaquin
EDUHSD Virtual Academy at Shenandoah (El Dorado)
iHigh Virtual Academy - San Diego

ELA 11
iHigh Virtual Academy - San Diego
Riverside Virtual

ELA data and results for high performing schools in the study are provided in Appendix B. A complete data set is available by request.

Mathematics

Most notable in mathematics was the general mathematics course at the iQ Academy in Los Angeles. Their mean value added was .776 standard deviations above the mean producing a minimum of .230 standard deviations above the mean with 95% confidence. The number of students was small, 11, and likely represents a single class with a single teacher. We note that due primarily to small individual classes in nearly all of the online algebra 2 courses, no single school performed at the distinguished level. The best school in this category was probably the Choice 2000 Online School in Riverside County with a mean value added of .396 standard deviations above the mean. The data covered only 18 students so the lower end of the 95% confidence interval extended down to -.120. Therefore this school does not meet our definition of distinguished.

Distinguished schools performing above average with 95% confidence

General Mathematics
Capistrano Connections Academy
iQ Academy LA

Algebra 1
CA Virtual Academy, Los Angeles
La Entrada Yorba Linda

Geometry
CA Virtual Academy, San Mateo
CA Virtual Academy, Los Angeles

Algebra 2
NONE

Mathematics data and results for high performing schools in the study are provided in Appendix C. A complete data set is available by request.

Other notable performances

By examining 90% confidence intervals we identified additional notable schools that were close to meeting our standard for distinction. We observed that a small number of schools performed well in multiple mathematics courses. Those schools were RAI Online Charter, performing notably in General Mathematics, Algebra 1, and Geometry. CA Virtual Academy at Los Angeles performed with distinction in both Algebra 1 and in Geometry. Capistrano Connections Academy performed with distinction in General Mathematics, but also performed notably in Geometry.
Limitations of the study

It must be emphasized that the primary objective of this initial research was to apply a value added approach to identify schools that indicated greater growth from one year to the next. The second phase of study would be to further investigate why this is the case. We would explore whether or not there was something in particular that these higher performing schools and programs were doing that resulted in greater growth over time. For this initial study, a simple value added model was applied that identified distinguished schools solely on the basis of the residuals of linear regression applied to cohorts of students that took the same pretest-posttest pair. It is becoming widely accepted among educators and researchers that proper identification of the true contribution of the educational experience to test performance must take into consideration a variety of additional factors. The data available for this analysis did not include additional factors and represents a limitation to the study. Another limitation to this study is the small number of students involved relative to the test-taking population of California. A few specific pretest-posttest cohorts that have large numbers statewide were excluded from this study because the number of students in our pool fell below 36. The online populations, while growing quite rapidly, still represent a very small percentage of the full population. As the online population continues to grow, opportunities for analysis that takes into consideration all test–pair cohorts and multiple years of performance will develop. These future studies will improve our ability to identify distinguished schools with greater certainty. Finally, this study is limited by the restraints of the current standardized testing system. Standardized tests are only one measure of student learning and represent a very narrow range of overall learning outcomes. In addition, their overall quality is limited to the types of subjects that lend themselves well to standardized tests.

Discussion and summary

We are witnessing tremendous growth in the number of public school students choosing to receive their education from authorized public online and hybrid schools. School leaders are being pressed to expand the number of authorized online schools. However, there is very little, if any research evaluating online schools or programs using value-added measures. The objective of this research was to illustrate how value-added methods can be used to identify online schools in California that perform with distinction compared with their counterparts. A simple value-added model was explained and applied to standardized testing results. The model measured educational growth differences at 32 schools in 8 subjects during the 2010-2011 academic year. The growth was based on California Standards Test data in successive years 2010 and 2011 for each student included in the study. The student pool consisted of those who were enrolled in one of 32 identified online or hybrid public schools at the time of the 2011 testing and whose corresponding test scores for 2010 were available.

This report included a review of the various value-added models. A brief discussion was provided pointing out the need for mindful policy development to avoid unintended consequences that may arise due to the control of non-educational variables in these models. A very basic value-added model was described and applied to the CST data. In this model no student characteristics were controlled. We avoided any distortion to the value-added calculations and did not account for variables such as free and reduced lunch densities, or other socio-economic or racial or any other factors. The underlying assumption in the selection of such a simple model is the idea that all students have similar capacities to learn. Despite our choice, we do not rule out that the control of some variables might need to occur to properly understand the relative quality of program outcomes. The selection of which variables to control requires significant discussion, well beyond the scope of this discussion. With the use of this model, the subsequent value-added results indicated the existence of distinguished online schools that perform above average with 95% statistical confidence in seven of the eight course categories analyzed.

Implications for further study

Educational effectiveness is multifaceted and any investigation into effectiveness should consider a multifaceted approach. Value-added measures only inform us about one dimension of the entire educational process (Condie, Lefgren, & Sims, 2014; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Konstantopoulos, 2014; Polikoff & Porter, 2014; Sanders, 2000) and should be viewed as one measure in a complex school or program
improvement process. This holistic approach to school improvement would support future research examining the quality of program level interventions, resources, supports and curriculum. Based on this rationale, the next phase of study would evaluate:

- The quality of program facilities and resources. Do students have an opportunity to collaborate and ask for help? If needed, are students provided additional resources and assistance such as access to online tutors or additional online content and practice for example?
- The quality of program content, knowledge and skill development (curriculum). Is the curriculum aligned with content area standards and to state standardized tests? Is it of sufficient rigor? Does it allow for teacher input and/or adaptation? Are extension activities built into the curriculum? Is there evidence of quality in online course design?
- The quality of program supports for students. Does the program offer extended-time learning opportunities? Are there math labs or tutorial sessions for struggling students for example? Is there a staffed help line or open lab hours? Are students provided multiple pathways for learning?
- The quality of program supports for teachers. Are teachers offered regular and consistent professional development in online teaching methods? Is peer coaching or mentoring integrated into professional practice?

In addition, because there is some evidence that increased certainty in estimated value-added scores have been shown over time (Cocoran, 2010; Ferrão & Couto, 2013), research using value-added measures should adopt a longitudinal approach. Future research would include applying the same value-added model of investigation to subsequent yearly cohorts.

Finally, online schools and programs offer a fairly unique opportunity in value-added investigations, in that results from these investigations can be linked directly to student and teacher behavioral data stored in server logs. Future studies using educational data mining combined with value-added measures may provide another avenue to further our knowledge and increase the confidence in value-added approaches to program evaluation.

Value added analysis of student performance over time can be a valuable tool when used appropriately. The results of our analysis should in no way be used as a judgment about the overall quality of any one school, course or teacher but rather as an initial large scale study, using aggregate data as part of a long term integrated analysis. Assessment and identification of best practices in online education is a growing national imperative and our intent was to focus on identifying high performing online schools and in so doing lay the foundation for further investigation into the what these identified “distinguished” schools are doing to promote long-term growth in student outcomes.

References


## Appendix A

### Technical data

Value Added Regression Model: \( \hat{y}_{jk} = \alpha_j + \beta_k x_{jk} + e_{jk} \)

#### English Language Arts Cohort Data

<table>
<thead>
<tr>
<th>Test Pair</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>Correlation</th>
<th>(N)</th>
<th>VA ST DEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>07 -&gt; 08</td>
<td>35.2028863</td>
<td>0.898920487</td>
<td>0.837081373</td>
<td>1349</td>
<td>33.63325091</td>
</tr>
<tr>
<td>08 -&gt; 09</td>
<td>75.3121477</td>
<td>0.794646791</td>
<td>0.819974315</td>
<td>1096</td>
<td>57.68691759</td>
</tr>
<tr>
<td>09 -&gt; 10</td>
<td>82.50142079</td>
<td>0.762449557</td>
<td>0.753896931</td>
<td>40</td>
<td>44.95626846</td>
</tr>
<tr>
<td>09 -&gt; 10</td>
<td>37.98750984</td>
<td>0.852815283</td>
<td>0.823073882</td>
<td>1298</td>
<td>32.44465709</td>
</tr>
<tr>
<td>10 -&gt; 10</td>
<td>136.8518591</td>
<td>0.588405325</td>
<td>0.601939595</td>
<td>37</td>
<td>45.9677917</td>
</tr>
<tr>
<td>09 -&gt; 11</td>
<td>-38.10792065</td>
<td>1.013138697</td>
<td>0.858618204</td>
<td>39</td>
<td>31.13760807</td>
</tr>
<tr>
<td>10 -&gt; 11</td>
<td>49.91924363</td>
<td>0.847915759</td>
<td>0.80259329</td>
<td>1178</td>
<td>36.1823348</td>
</tr>
<tr>
<td>11 -&gt; 11</td>
<td>85.21475599</td>
<td>0.762701545</td>
<td>0.704078496</td>
<td>48</td>
<td>40.31095084</td>
</tr>
</tbody>
</table>

*Note.* VA ST DEV = standard deviation of cohort scaled score residuals. Example: 07-> 08 pair is ELA grade 7 pretest, ELA grad 8 posttest pair.

#### Mathematics cohort data

<table>
<thead>
<tr>
<th>Test Pair</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>Correlation</th>
<th>(N)</th>
<th>VA ST DEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-&gt;1</td>
<td>86.33706271</td>
<td>0.731949056</td>
<td>0.683467932</td>
<td>255</td>
<td>38.81331348</td>
</tr>
<tr>
<td>1-&gt;1</td>
<td>65.6262479</td>
<td>0.802543444</td>
<td>0.763661159</td>
<td>190</td>
<td>32.07149261</td>
</tr>
<tr>
<td>3-&gt;1</td>
<td>127.7321388</td>
<td>0.68344329</td>
<td>0.618302607</td>
<td>92</td>
<td>46.80170421</td>
</tr>
<tr>
<td>0-&gt;3</td>
<td>75.91702932</td>
<td>0.628227061</td>
<td>0.684321574</td>
<td>1031</td>
<td>42.57601403</td>
</tr>
<tr>
<td>1-&gt;3</td>
<td>130.7451863</td>
<td>0.443840181</td>
<td>0.597067193</td>
<td>447</td>
<td>33.55203361</td>
</tr>
<tr>
<td>3-&gt;3</td>
<td>99.34208976</td>
<td>0.656159185</td>
<td>0.691095793</td>
<td>557</td>
<td>32.140072</td>
</tr>
<tr>
<td>5-&gt;3</td>
<td>129.1694387</td>
<td>0.59520695</td>
<td>0.597078508</td>
<td>46</td>
<td>37.98014106</td>
</tr>
<tr>
<td>1-&gt;5</td>
<td>103.8829867</td>
<td>0.531670478</td>
<td>0.713017217</td>
<td>129</td>
<td>35.48511869</td>
</tr>
<tr>
<td>3-&gt;5</td>
<td>123.9689371</td>
<td>0.520891358</td>
<td>0.62430379</td>
<td>827</td>
<td>39.9145496</td>
</tr>
<tr>
<td>5-&gt;5</td>
<td>84.37919208</td>
<td>0.713123308</td>
<td>0.64708855</td>
<td>108</td>
<td>31.80480223</td>
</tr>
<tr>
<td>3-&gt;7</td>
<td>-32.84972536</td>
<td>0.911098088</td>
<td>0.797665759</td>
<td>36</td>
<td>46.07957427</td>
</tr>
<tr>
<td>5-&gt;7</td>
<td>72.81724677</td>
<td>0.650672789</td>
<td>0.679836044</td>
<td>429</td>
<td>36.0494139</td>
</tr>
</tbody>
</table>

*Note.* Test Codes: 0 = 7th Grade Math; 1 = 8th-9th General Math; 3 = Algebra; 1, 5 = Geometry; 7 = Algebra 2.
Appendix B

English Language Arts

Course: ELA Grade 8

<table>
<thead>
<tr>
<th>Course</th>
<th>ELA 8</th>
<th>95% Confidence</th>
<th>90% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean VA</td>
<td>VA Min</td>
</tr>
<tr>
<td>CA Virtual Academy, San Diego</td>
<td>234</td>
<td>0.135</td>
<td>0.027*</td>
</tr>
<tr>
<td>Capistrano Connections Academy</td>
<td>108</td>
<td>0.147</td>
<td>-0.011</td>
</tr>
</tbody>
</table>

*Note.* Per California CDE policy data is deleted when N is below 11 students. VA = value added, VA Min = lower limited of the confidence interval, N = student count. *indicates entirely positive 95% confidence intervals; **indicates entirely positive 90% confidence intervals.

Course: ELA Grade 9

<table>
<thead>
<tr>
<th>Course</th>
<th>ELA 9</th>
<th>95% Confidence</th>
<th>90% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean VA</td>
<td>VA Min</td>
</tr>
<tr>
<td>CA Virtual Academy, Kings</td>
<td>38</td>
<td>0.286</td>
<td>0.019*</td>
</tr>
<tr>
<td>CA Virtual Academy, Los Angeles</td>
<td>258</td>
<td>0.197</td>
<td>0.094*</td>
</tr>
<tr>
<td>CA Virtual Academy, San Diego</td>
<td>156</td>
<td>0.183</td>
<td>0.051*</td>
</tr>
</tbody>
</table>

Course: ELA Grade 10

<table>
<thead>
<tr>
<th>Course</th>
<th>ELA 10</th>
<th>95% Confidence</th>
<th>90% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean VA</td>
<td>VA Min</td>
</tr>
<tr>
<td>CA Virtual Academy, Kings</td>
<td>39</td>
<td>0.302</td>
<td>0.038*</td>
</tr>
<tr>
<td>CA Virtual Academy, Los Angeles</td>
<td>239</td>
<td>0.195</td>
<td>0.088*</td>
</tr>
<tr>
<td>CA Virtual Academy, San Joaquin</td>
<td>33</td>
<td>0.338</td>
<td>0.051*</td>
</tr>
<tr>
<td>Capistrano Connections Academy</td>
<td>113</td>
<td>0.238</td>
<td>0.082*</td>
</tr>
<tr>
<td>EDUHSD Virtual Academy at Shenandoah (El Dorado)</td>
<td>29</td>
<td>0.360</td>
<td>0.045*</td>
</tr>
<tr>
<td>iHigh Virtual Academy - San Diego</td>
<td>18</td>
<td>0.765</td>
<td>0.355*</td>
</tr>
</tbody>
</table>

Course: ELA Grade 11

<table>
<thead>
<tr>
<th>Course</th>
<th>ELA 11</th>
<th>95% Confidence</th>
<th>90% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean VA</td>
<td>VA Min</td>
</tr>
<tr>
<td>CA Virtual Academy, Los Angeles</td>
<td>235</td>
<td>0.085</td>
<td>-0.023</td>
</tr>
<tr>
<td>CA Virtual Academy, San Diego</td>
<td>148</td>
<td>0.123</td>
<td>-0.013</td>
</tr>
<tr>
<td>CA Virtual Academy, San Mateo</td>
<td>77</td>
<td>0.163</td>
<td>-0.025</td>
</tr>
<tr>
<td>iHigh Virtual Academy - San Diego</td>
<td>16</td>
<td>0.525</td>
<td>0.087*</td>
</tr>
<tr>
<td>Riverside Virtual</td>
<td>17</td>
<td>0.603</td>
<td>0.180*</td>
</tr>
</tbody>
</table>
Appendix C

Mathematics

**Course: General Mathematics**

<table>
<thead>
<tr>
<th></th>
<th>General math</th>
<th>95% Confidence</th>
<th>90% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean VA</td>
<td>Mean Min</td>
</tr>
<tr>
<td>Capistrano Connections Academy</td>
<td>96</td>
<td>0.224</td>
<td>0.056*</td>
</tr>
<tr>
<td>iQ Academy LA</td>
<td>11</td>
<td>0.776</td>
<td>0.230*</td>
</tr>
<tr>
<td>RAI Online Charter - San Diego</td>
<td>23</td>
<td>0.322</td>
<td>-0.036</td>
</tr>
</tbody>
</table>

**Course: Algebra 1**

<table>
<thead>
<tr>
<th></th>
<th>Algebra 1</th>
<th>95% Confidence</th>
<th>90% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean VA</td>
<td>Mean Minimum</td>
</tr>
<tr>
<td>CA Virtual Academy, Kings</td>
<td>79</td>
<td>-0.124</td>
<td>-0.310</td>
</tr>
<tr>
<td>CA Virtual Academy, Los Angeles</td>
<td>581</td>
<td>0.115</td>
<td>0.047*</td>
</tr>
<tr>
<td>La Entrada Yorba Linda</td>
<td>11</td>
<td>0.632</td>
<td>0.086*</td>
</tr>
</tbody>
</table>

**Course: Geometry**

<table>
<thead>
<tr>
<th></th>
<th>Geometry</th>
<th>95% Confidence</th>
<th>90% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean VA</td>
<td>Mean Minimum</td>
</tr>
<tr>
<td>CA Virtual Academy, Los Angeles</td>
<td>264</td>
<td>0.151</td>
<td>0.050*</td>
</tr>
<tr>
<td>CA Virtual Academy, San Mateo</td>
<td>84</td>
<td>0.250</td>
<td>0.070*</td>
</tr>
<tr>
<td>CA Virtual Academy/Jamestown - Tuolumne</td>
<td>12</td>
<td>0.406</td>
<td>-0.112</td>
</tr>
<tr>
<td>Capistrano Connections Academy</td>
<td>60</td>
<td>0.191</td>
<td>-0.022</td>
</tr>
<tr>
<td>RAI Online Charter - San Diego</td>
<td>11</td>
<td>0.530</td>
<td>-0.017</td>
</tr>
</tbody>
</table>
Instructional Design and Professional Informal Learning: Practices, Tensions, and Ironies

Stephen C. Yanchar* and Melissa N. Hawkley
Instructional Psychology and Technology, Brigham Young University // stephen_yanchar@byu.edu // mhawkley@gmail.com

*Corresponding author

(Submitted September 6, 2014; Revised February 13, 2015; Accepted March 12, 2015)

ABSTRACT
This qualitative study explored the nature of informal learning in professional instructional designers’ everyday work activities. Based on intensive interviews with six full-time practitioners, and using a hermeneutic form of data analysis, this study produced seven themes concerning the practices, tensions, and ironies associated with this largely unexplored aspect of instructional design. Specific themes concerned the resources, significance, and outcomes of informal learning, in addition to professional tensions and ironies (e.g., the observation that informal learning simultaneously fosters professional survival and professional vulnerability). Implications are discussed, such as how informal learning among professionals can be facilitated via organizational support and more practically-oriented scholarly work, as well as designers’ own efforts to pursue informal learning in the midst of practice.

Keywords
Instructional design, Informal learning, Design practice

Introduction
How do instructional designers learn their trade? A number of studies in recent years have addressed this question, offering insight into what might be generally thought of as the process of becoming a designer in the field (e.g., Ertmer et al., 2008; Hardré, Ge, & Thomas, 2006). In some ways, it seems clear that instructional designers learn their trade the same way as practitioners in other fields—through formal training as a novice and a variety of experiences afterward. As professionals dedicated to the effective instruction of others, instructional designers must surely recognize their own need for continued professional development. In this sense, it seems reasonable to assume that instructional designers would be dedicated to learning in two broad senses—the learning of others as well as their own.

A good deal of literature on becoming an instructional designer has focused on efforts to improve formal training. A common trend in this regard concerns the inculcation of design skills and an accompanying shift beyond process models assumed to guide designers through the process of creating learning experiences (e.g., Rowland, Fixl, & Yung, 1992; Visscher-Voerman, Kuiper, & Verhagen, 2007). Work by Rowland and colleagues offers a good example of this perspective. As Rowland, et al. (1992) suggested, adequate training should result in skills that facilitate what instructional designers do in workplace settings. In particular, they advocated an apprentice model that entails three elements: learning in context, modeling of expertise, and reflection. In a related commentary, Rowland, Parra, and Basnet (1994) advocated a “creative” view of design and enumerated a series of methods for training instructional designers based on educational practices in other design fields such as engineering and architecture. Their recommendations included training methods such as case studies, competitions, internships, design studios, and expert demonstrations.

Other literature related to becoming a designer concerns expert performance (e.g., Ertmer et al., 2008; Hardré, Ge, & Thomas, 2006; Kirschner, Carr, van Merriënboer, & Sloep, 2002; Rowland et al., 1992), demonstrating the importance of prior design experience, an ability to analyze complex design problems, and systematic strategies for dealing with that complexity. Others suggest that training should be enhanced by coverage of applied knowledge and skills, as in design studio approaches (e.g., Clinton & Reiber, 2010). Thus, practical aspects of instructional design work appear to be increasingly emphasized in discussions of professional training.
Given this emphasis on practical aspects of work in the field, it is noteworthy that little attention has been paid to instructional designers’ efforts to expand their capabilities while on the job. For example, discussions of expert-novice differences, theoretical vs. practical knowledge, and improved training methods all focus on formalized educational experiences that typically occur in structured settings and, for the most part, away from fields of actual practice. There is in the literature no substantive examination of workplace experiences through which practicing designers become increasingly skilled at their craft. Practical experiences that occur during training, such as those had in internships, are surely interesting in this regard; but the related question of how instructional designers continue to learn and develop professionally, long after formal training and internships have been completed, stands as a unique and under-researched area of scholarship in the field.

In conducting this study, we focused on what scholars in the adult education literature have termed informal learning in the workplace. Such learning is primarily concerned with professional development that occurs outside of formalized curricular structures. Following the conventions typical within the adult education literature (e.g., Merriam, Caffarella, & Baumgartner, 2007), the label informal learning has to do with the ordinary, unstructured means by which practitioners cope with everyday tasks and become more capable of completing work responsibilities. This is learning that allows not only for the successful completion of projects, but also for the progressive refinement of skills over time. In the adult education literature, formal workplace learning is often contrasted with two other types of learning, namely, formal and nonformal. While the former of these is associated with highly-structured, instructional settings with a formal curriculum (e.g., public school education), the latter exists outside of formal educational systems (e.g., community workshops sponsored by a local library or university extension service), though still includes structured learning experiences with an instructor, curriculum, and so on. In contrast to formal and nonformal learning experiences, then, informal learning refers to one’s personal effort to achieve greater competence in one’s profession, without curricular support.

As others have observed, informal learning plays a crucial role in workplace effectiveness. One set of researchers, for example, estimated that well over half of learning that occurs in work settings is informal in nature (Merriam et al., 2007). Other analyses have suggested that informal learning is vital to many professions (Cheetham & Chivers, 2001, 2005), with specific studies conducted in fields such as teaching (Lohman, 2006), engineering (Reardon, 2004), information technology (Jubas, Butterwick, Zhu, & Liptrot, 2006), and global leadership (Cseh, Davis, & Khilji, 2013). Across this growing literature there is evidence of focused concern with this unique form of learning. While we are aware of no published data or estimates regarding the amount of informal learning in instructional design and technology as a field, we find it reasonable to assume that the degree of informal learning required of practitioners in this profession is at least equal to, if not greater than, the degree of such learning in other professions.

In what follows, then, we offer a qualitative examination of informal learning among practicing instructional designers. We do so with two purposes in mind: first, to extend the emerging body of scholarship on everyday instructional design practices; and second, to suggest what these findings may imply regarding continued professional development of practitioners in the field. Thus, we seek to provide some insight regarding questions such as: What is the nature of informal learning for instructional designers? What is the practical significance of informal learning in the field? And what are some of the realities and challenges faced by designers as they seek to augment their skills?

**Method**

**Study overview**

We investigated informal learning in instructional design practice with a research strategy that emphasized participants’ reflections on everyday design experiences. More specifically, our strategy synthesized elements of hermeneutic (Kvale & Brinkmann, 2009), phenomenological (Giorgi & Giorgi, 2003) and ethnographic (Spradley, 1979) qualitative approaches that, taken together, provided a basis for exploring the meaning of everyday work activities in context, with particular emphasis on designers’ efforts to gain skills, solve problems, and generally refine their craft.
Participants

The study included six participants, all of whom were employees at an instructional design center at a major university (see participant information in Table 1). Five participants were primarily engaged in instructional design work, although one specialized in educational videography, another worked primarily as an artist as well as designer, and another had recently assumed development as well as design responsibilities at the center. The sixth participant was trained in design and usability, and now worked primarily as a trainer, instructing (or creating instruction for) faculty on products created in the center.

Table 1. Participant information

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Total years experience</th>
<th>Role at the center</th>
<th>Highest degree in the field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alison</td>
<td>Female</td>
<td>25</td>
<td>Designer</td>
<td>MS</td>
</tr>
<tr>
<td>Sylvia</td>
<td>Female</td>
<td>4.5</td>
<td>Designer</td>
<td>PhD</td>
</tr>
<tr>
<td>Lisa</td>
<td>Female</td>
<td>2</td>
<td>Trainer/Training Designer</td>
<td>MS</td>
</tr>
<tr>
<td>Martin</td>
<td>Male</td>
<td>22</td>
<td>Artist/Designer</td>
<td>MS</td>
</tr>
<tr>
<td>William</td>
<td>Male</td>
<td>18</td>
<td>Educational Videographer</td>
<td>MS</td>
</tr>
<tr>
<td>Greg</td>
<td>Male</td>
<td>12</td>
<td>Designer/Developer</td>
<td>PhD</td>
</tr>
</tbody>
</table>

Interview procedure

We conducted two semi-structured interviews per participant. This procedure enabled us to probe into our phenomena of interest fairly intensively. Most interviews lasted about an hour. The first author conducted all interviews and was often assisted by the second author, who also asked questions during the interview sessions. In the first interview, we queried broadly into participants’ background, everyday work life, and general thoughts about learning at work. In the second interview, we asked more targeted questions regarding specific informal learning work experiences and participants’ views of them. At the close of each interview we invited participants to share any additional thoughts about the topic or our interview process.

Data analysis

Through the transcription process we engaged in light editing to make some of the statements more readable (e.g., eliminating the sound “uh” between words). We were careful to edit in ways that did not alter participants’ meanings or eliminate important events (e.g., a long, thoughtful pause). Our analysis involved a hermeneutic form of data condensation and thematic analysis (Kvale & Brinkmann, 2009)—augmented by techniques from descriptive phenomenology (Giorgi & Giorgi, 2003)—that facilitated our effort to produce an insightful account of the activities and phenomena related to workplace learning among instructional designers. In brief, our data analytic process entailed reading the transcripts to gain a sense of the whole, creating a list of possible themes, narrowing the list of themes down to those best supported by the data and most capable of providing insight, and finally, identifying quotes that best illustrate the selected themes.

Trustworthiness

We attempted to treat the data as fairly as possible by utilizing well-accepted qualitative standards of trustworthiness (Lincoln & Guba, 1985). These procedures included peer debriefing, member checking, negative case analysis, and progressive subjectivity checks. As our member checks suggested, no participants expressed concern about their quotes or the themes we generated. Two participants requested minor revisions to the wording of quotes, to clarify a point or omit awkward phrasing, which did not alter their meanings. Regarding progressive subjectivity checks, we kept a record of our assumptions, values, and details regarding our research process. Regarding peer debriefing, we sought feedback from a qualified peer who suggested that our research design was appropriate for our subject matter and met acceptable standards for trustworthiness. Regarding negative case analysis, we sought out contradictions and counterexamples to emerging themes as we analyzed data. While participants were not monolithic in their views, they did share significant agreement on many issues; the interesting differences in their views are reflected in the themes and quotes we presented.
Findings

We present our findings in terms of: (a) three local patterns, (b) two significant tensions, and (c) two ironies that showed up in the narratives related by our participants (see Table 2).

<table>
<thead>
<tr>
<th>Three local patterns</th>
<th>Resources of informal learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Significance of informal learning</td>
</tr>
<tr>
<td></td>
<td>Results of informal learning</td>
</tr>
<tr>
<td>Two tensions</td>
<td>Continuity and change</td>
</tr>
<tr>
<td></td>
<td>Satisfactions and challenges</td>
</tr>
<tr>
<td>Two ironies</td>
<td>Ubiquity and invisibility</td>
</tr>
<tr>
<td></td>
<td>Survival and vulnerability</td>
</tr>
</tbody>
</table>

Three local patterns

The three themes we describe here as local patterns are not presented as lawful or statistical generalizations, but rather as local patterns, in that they provide a glimpse into the phenomenon of informal learning at this organization and offer what has been termed “transferable” insight (see Lincoln & Guba, 1985).

Resources of informal learning

Although participants’ ability to articulate aspects of their professional practices varied, they discussed resources that facilitate their informal learning with relative ease. Examples of such resources include conversations with other designers or team members (e.g., designers’ meetings), observing others at work, examining what other people have produced, usability data and user concerns, subject matter experts, and many different resources on the internet that provide useful ideas. A typical statement in this respect was as follows:

Sylvia: you can always Google and see what you can find out. From other people you can follow design sites and usability sites and things like that and read the literature of what people are talking about to stay up on top of things… Interviewer: You mentioned looking at other people’s work for examples, right? Sylvia: Yeah, so you can go out and play with other programs, you can go and ask questions, you can be reading the current literature, you can be doing a lot of things on the internet, but I also think no one on the internet quite knows you and your users and your product like the people around you. So if you’re not talking to your users and you’re not talking to your colleagues, you’re really missing out on a lot of feedback and a lot of opportunities to really grow.

Experience and trial and error were also commonly mentioned as sources of learning:

William: We were starting on the first script, we had like, I think there were 10 scripts that we had to shoot in this week long period, 10 different segments. And after shooting the first one, I realized what she wanted. And that first one was just a few pages; it shouldn’t have taken as long as it did to shoot. And I said okay we’ve got to make a change here. So then we set up that camera to go over the shoulder and the next script took half as long as that first one. And then as we got through that process, we recognized what she was really after, and then we were able to turn it over quicker and quicker.

In a related discussion on how practical wisdom develops, William offered this concise response: “A lot of errors, a lot of mistakes.” In this sense, it might be concluded that if mistakes are inevitable, which they appear to be, then practitioners should be willing to view them as learning opportunities and adjust in ways that promote future success.

Significance of informal learning

Two participants saw informal learning as an inescapable part of professional work, in addition to life in general, suggesting the importance of continual adjustment to contextual circumstances and the practical wisdom it produces.
In this sense, learning was linked tightly to professional survival for these designers. When speaking of the need for continual learning, Greg stated: “And then just the nature of life; there’s always things that come up that you need to learn.”

Sylvia commented on the urgency of informal learning, at least at times:

…and you had to learn. It wasn’t like here do this project and get back to us in a few months; it was you need to start giving us stuff now because we need to have it programmed in a year. So it was just fast.

Lisa, as the others we interviewed, suggested that she couldn’t succeed in her work without continual informal learning:

Lisa: if we’re the support organization we have to understand continually how it works. Interviewer: So you couldn’t do your job without that kind of …? Lisa: No, no I need to keep up on top of it or I can’t.

Participants also discussed how much of what they have learned to do in their positions can’t be taught in formalized class settings. As Lisa remarked:

I think you can provide a structure and some resources for people to do things but really I feel like a lot of the deep learning happens when you’re actually doing stuff. I’ve read through something before and thought I understood it, but when I actually applied it, I had an “aha” moment where the learning actually happened.

The results of informal learning

Most informal learning situations described by our participants—such as learning new software, becoming acquainted with new co-workers, and orienting to new projects—were not particularly momentous, but still happened regularly and needed to be handled in some way … Informal learning in these cases leads to a gradual deepening of designers’ skills. However, more dramatic shifts in perspective were also described by our participants as a significant result of informal learning. That is, informal learning often entailed new ways of looking at various aspects of projects and new ways of being involved in their work. For example, participants discussed how they were able to solve problems as a result of insights or changes in perspective. Lisa suggested that this is an important aspect of her work:

Interviewer: Does that seem like learning to you or does it seem like something else? Lisa: In a lot of ways I do feel like it’s learning, just because it’s something new, it’s a new way of thinking about it, it’s meeting an expectation in a different way.

As an example, Sylvia commented on how her view of good design has changed through experience. Early in her career she believed that there was a single best design in any circumstance, but later realized that there may very well be no optimal approach in a given situation and that getting it “right” means that user needs are met, which can happen in more than one way, including ways that violate traditional design principles. Over time, her ability and confidence to design in this way has increased.

Two tensions

Continuity and change

As important as informal learning and its accompanying change showed itself to be for our participants, they did not necessarily embrace each opportunity to learn something new. Indeed, at times they actually avoided a learning encounter, typically for strategic purposes, and relied on existing skill. In this sense, informal learning, when it did take place, was purposive and selective—based on a designer’s time, energy, and perception of the significance of the opportunity to learn. One participant noted that learning when to pursue or avoid such opportunities was itself an important achievement.
Interviewer: So that itself was a kind of learning—learning to know when to dump it off on someone else and when to take the challenge. Alison: Yes...economies of time say let the expert do it when it’s a rare thing. But other things, especially with tools, you have to learn them; that just takes time.

With regard to this topic, William mentioned a former co-worker who constantly explored new technologies. William, in contrast, preferred a degree of continuity over such constant exploration, using familiar tools he considered to be effective.

On the other hand, there are times when learning cannot, or should not, be avoided. As Alison recounted: “I was designing instruction for consultants and an actual artifact that they would be using. . . I couldn’t pass that off to anyone. That was something I had to learn.”

Similarly, Martin noted that learning is sometimes what accelerates a project’s completion:

We know we can make it look good enough by scratch, but that’s way too much work. Let’s try to solve this problem in a different way. And so the learning was like finding out what the other options are, picking one that works the best and then we had to jump into that particular tool and figure out—ok, yeah we know this tool can do it, but what really do we have to do to make it happen?

Martin also indicated that there are times when one’s skills and perspective are adequate to the task, and additional learning beyond what is truly essential is often not worth the effort. As he stated: “I have to balance the fun and excitement and satisfaction of continually learning versus, I’ve got to get this batch of stuff done and so learn as much as you can to get that stuff done.”

Thus, whether or not designers in our study chose to learn seemed to depend upon the significance of the encounter, balanced against the other situational demands. At times, the change brought by informal learning was deemed worthwhile or even crucial; at other times, prior experience, skill, and know-how were adequate to complete a task. A tension was created, in this sense, by circumstances that strained an instructional designer’s abilities and created a choice between continuity and change.

**Satisfactions and challenges**

Several of our participants worked closely with subject matter experts in a given area, which they described as an interesting and gratifying part of their work. Another participant emphasized the satisfaction that comes from helping others in the course of her duties. But for all of the participants in our study, learning itself—including being stretched and taking on new challenges—was viewed as a satisfying part of work. Sylvia expressed this sentiment in the strongest terms:

Sylvia: To me I don’t think I would stay in a job if I wasn’t growing. Other people have talked to me about other jobs and things like that and I always say, if I’m ever at a point where I’m not growing, then, yeah, I would consider other things.

For our participants, however, the satisfying aspects of informal learning are balanced against the frustrations of the work, including situations in which committing time and energy to learn something new seems to delay progress. As Sylvia succinctly concluded: “So it’s good—learning—but sometimes in that moment there’s a lot of frustration and a lot of hard times.”

The tension between frustration and satisfaction came out most clearly in the following quote about informal learning:

Greg: There are days where that’s what energizes me and there are days where it overwhelms and consumes you, like, “I’ll never be held to keep up! Ahh! Can I retire? No, I’ve got to learn.”
Two ironies

Ubiquity and invisibility

The course of our participants’ work activities might be best described as uneven and bumpy, interrupted by the need to adjust to what a situation has presented. In our participants’ experience, these situations, and the learning that often ensues, are inevitable. As one participant stated:

Greg: Each new project presents its own little challenges—some require new learning, some not as much, although I don’t think you ever want to take things for granted—that you know everything you need to know.

Some challenges are fairly modest, such as learning new features of software. Others are significant and multifaceted, as when one participant took on a new role within the organization:

Greg: I’m in a position now where I had to take over development responsibilities for somebody on our team that left … I was suddenly thrust into taking on additional job responsibilities where I had to try to get up to speed in things like Drupal and Wordpress and, just how to how to do some things related to servers and keeping things up and running.

Early in the interviews, our participants had some difficulty remembering obvious or common learning experiences, beyond those involving new technology and course subject matter; other aspects of learning were less apparent to them. As our participants reflected more on their experiences, however, they began to see other aspects of their work as learning, such as learning how to solve problems, create course designs, work effectively with co-workers, and cope with complex, rapidly-changing workloads. Through our interviews it was clear that shifts in their skills and perspectives were intertwined with everyday work activities and that learning itself is part of normal work life. Sylvia summarized the everydayness of informal learning in this way:

Sylvia: I think all the time you’re informally learning. I don’t know. Part of me thinks that’s what we are as part of life and living is we’re always learning from our situations and surroundings and working to improve or working to do something. So at [our organization] I feel like every day all the time is informal learning.

Initially, Martin equated learning with explicit efforts to gain skill in new subject matter or unfamiliar technologies. Martin estimated that learning of this sort took up a fairly small amount of his work time. In a strict sense, Martin’s estimate counts as a negative case regarding the commonality of informal learning in instructional design/development work. Through further discussions with Martin, however, it became clear that, in his experience, informal learning occurs at virtually all stages of the design/development process, and particularly as his team sought to understand client needs and find ways to produce the best learning environment possible, under certain circumstances. As Martin clarified:

…a large part of the production is coming up with new ideas and figuring out a new way to communicate something. And so, in a sense, that’s kind of learning the creative process; we’re creating something that didn’t exist, that we need to figure out how to solve this problem, and that’s kind of, I don’t know, that’s kind of learning and that’s part of the mix of production. … For me that’s just part of what needs to happen. … And so, I have to learn what that environment is like and what’s going to make a difference for the students and what you want to include in it and stuff like that. And so, in a sense I’m learning a lot about you and your students.

William suggested that gaining an understanding of client needs and personalities was an important part of video production. Given that clients are involved in every project, this amounts to a significant aspect of William’s work:

That’s part of the process. And so learning to deal with clients is really what I do more, unfortunately, more than the setting up the camera and the writing. It’s communicating with clients that is the most important thing.

In this sense, the work of designing requires learning at many levels. One of those levels would seem to be design itself, which is, in this sense, learning about a situation—some problem or opportunity—and working out a feasible plan to produce what is needed.
Notwithstanding the ubiquity of informal learning, however, our participants typically didn’t experience it as learning *per se* in the moment. Thus, informal learning often happens implicitly or semi-consciously. As Lisa suggested:

Well, I feel like it’s something that I don’t really think about a lot. I think it just kind of happens and all of a sudden I’ve learned something new. It’s not something I really consciously think about a put a lot of effort into, “Oh I learned something today…”

Later in the interview, Lisa commented, “…we don’t think about it as learning.”

There appears to be an irony, then, in the complex nature of informal learning, which was, for our participants, both ubiquitous and experientially invisible. Informal learning enabled designers to continuously adapt to their work conditions, yet such learning was not commonly viewed as learning *per se*, and was perhaps so integral to their work that it was not seen as anything but design itself. Thus, what is highly important and “close” experientially was recognized and reflected on the least. In this sense, it appears that design work inherently involves a degree of innovative learning, such that this learning is not viewed as *learning* at all, but an ordinary aspect of design work.

**Survival and vulnerability**

As we suggested above, learning about (and adapting to) the unique demands of a given project was central to our participants’ experiences as instructional designers, in addition to whatever other learning may be required, such as becoming proficient in relevant technology, managing relationships of all sorts, and becoming acquainted with concepts and models that facilitate work on a given project. At least for our participants, then, workplace learning is considered a means of what might be considered professional survival.

On the other hand, such informal learning that allows for survival can become stressful or even overwhelming and lead to feelings of professional vulnerability. When queried about the range of informal learning opportunities in his experience, Greg distinguished between relatively manageable learning encounters and high-stakes situations that require substantial time and effort. In his words:

Yeah the learning itself, well, generally, a lot of it’s “just in time” kind of stuff; but when they first said [Greg], we want you to know that [another employee] is leaving—we can’t afford to replace him and so you’re taking over his responsibilities—that was a decision point where I had to decide, okay what am I going to do? And that’s where I kind of went back to my desk and had a few other things, but this was mulling over in my mind, how on earth am I going to get to where I need to be to even survive, let alone thrive, with this initial responsibility?

Through subsequent interviews, Greg reported that he was able to “survive” in the months following this revelation and successfully adapt to his new responsibilities; but not without considerable effort and some anxiety on his part.

Lisa also mentioned the stress that can accompany workplace learning, especially when confronted with novel situations that she didn’t feel prepared to address. As she engaged in the implementation of faculty training, she was sometimes asked questions about the technology that she couldn’t immediately answer. Her thoughts regarding such situations were as follows:

It’s hard because I’m supposed to be the one who knows stuff that they have questions about, and when I don’t it’s really hard for me to say, “ok.” But I’ve found if I can just say, “You know what, give me just a couple of minutes,” or “Let me get right back to you,” and then I can go back to my computer and back to my own space and usually within a few minutes I can figure out what the issue is. But I have a hard time right in the very moment. It’s really hard.

Alison raised a concern that was similar in some ways, suggesting that her perception of how others will view her queries when designing can impede her efforts to seek help.
Alison: Part of it sometimes is you don’t want to go to a colleague and say, “Help me with this” for fear they are going to say, “You don’t know that?” [laughing] … So there is always this inhibition, there is always the danger of exposing your stupidity—your lack of knowledge.

These instances suggest that learning on the job, as important as it may be, often comes with some degree of risk that may leave a designer feeling professionally vulnerable—that is, vulnerable to stress and frustration, to a loss of respect among co-workers, or possibly even dismissal if one’s skills and potential to grow professionally are deemed insufficient. There appears to be an irony, then, in that the very activity which enables professional survival can, in some circumstances, lead to vulnerabilities or create an impediment to professional success.

Discussion

The findings of this study differ from other investigations of design practice by exploring actual learning activities of practitioners rather than more traditional topics such as expert novice differences and expert problem solving (e.g., Ertmer et al., 2008; Kirschner, et al., 2002). Our participants’ accounts pointed to the ubiquity of informal learning as a means of everyday coping in the workplace. There is a sense, however, in which this reliance on creative exploration and adjustment in context might be viewed somewhat negatively, particularly if one views instructional design as a predictable process guided by a form of technical rationality. While this position is not endorsed by all commentators in the field, it does represent a distinct historical way of viewing design practice (Smith & Boling, 2009). As should be clear, our participants relied heavily on their own personal ways of solving instructional problems and, as suggested in our data, learning how to navigate the difficulties that arise when seeking to balance quality and feasibility. If informal learning is as significant for instructional designers elsewhere as it was for our participants, then informal learning efforts by designers should not be viewed as nontechnical sloppiness, but rather as an essential design activity. Design practice that entails such learning may be challenging at times, but is essential for professional survival. Moreover, the shifts in perspective that are part of informal learning may foster some of the most significant advances in professional development. Thus, the irony that informal learning leads to both survival and vulnerability appears to be, in some sense, unavoidable.

The seven themes we have presented here hold several implications for instructional design. First, we suggest that, to whatever degree possible, instructional designers’ informal learning efforts should be explicitly supported by their organizations. Some amount of time, access to knowledge, and financial resources could be allocated as a sort of investment by which designers could expand their perspectives and skills, not only as part of formal professional development, but also in the midst of projects in which informal learning can be leveraged to increase productivity and quality. More specifically, such support could entail project deadlines that anticipate the time and space needed for designer learning, work time dedicated to informal discussion groups for the exchange of ideas (not directed or supervised by management), access to helpful literatures through online sources or an institutional library, the creation of proprietary case studies that document learning in the process of design, and related resources. Moreover, an organizational emphasis on learning and creative solutions to instructional problems (e.g., rather than emphasizing templatized work, at least in many cases) could demonstrate how an organization values this important aspect of design, possibly accompanied by professional reward structures that incentivize effective learning in these ways. Also, in this regard, conceptual models of informal learning that facilitate the process of becoming a learner-designer could be used as job aids or integrated into formal instructional design training. What such models would look like remains to be determined, but they could offer a sense of best informal learning practices, especially for novices in the field. Future research that explores best practices in this area could inform the design of such models.

Moreover, support for designers could entail the development of an organizational culture in which professional learning is valued and nurtured. For example, one of the most interesting comments was offered by Alison, as she expressed concern about exposing what might be considered professional inadequacies as she sought help from coworkers. Although we have no direct evidence that others would see informal learning in this light, it seems reasonable to assume that exposing such limitations may leave designers feeling professionally at risk. To facilitate informal learning, it would be beneficial for workers to be provided with an atmosphere conducive this kind of collegial interaction. Future research could probe more deeply into the vulnerability that designers may perceive in such situations, as well as how they cope with the professional challenges it brings. Moreover, alternative inquiry approaches such as multi-site surveys could offer a broader view of this phenomenon—as well as others reported in this study—and greater insight regarding how frequently and intensely it is experienced.
A second implication of this inquiry is that the most promising scholarly way to support instructional design practice will not be through design formalisms, such as process models and prescriptive theories expected to guide practice to a predictable outcome. If design entails a considerable learning component, as it seems to for our participants, then best practices will not demand adherence to a process or theoretical orthodoxy. Moreover, when queried, participants suggested that the progress they achieve through this type of learning is not made possible in the curricula of formal training. While such training surely helps prepare the uninitiated for beginning ventures in the field, its emphasis on academic formalisms cannot offer specific guidance regarding complex and rapidly changing work situations that call for flexibility, resourcefulness, and contextual problem solving. On the other hand, there is some evidence that such formalisms can be helpful as part of a conceptual “tool kit” (Yanchar et al, 2010). Thus, learning when and how to adapt various conceptual tools to certain applications fits within the design-by-learning account we have offered. Moreover, flexible models and principles that offer adaptable concepts to be developed by practitioners in actual practice and tailored to specific contexts would offer a desirable goal for future scholarly work. Resources of this type are not in wide supply, but have been discussed in the literature (Yanchar & Faulconer, 2011).

A third implication of this inquiry concerns instructional designers’ willingness to engage informal learning opportunities in productive ways. It is interesting that many informal learning efforts are experientially “invisible,” despite their ubiquity in the work of instructional designers. Theorists in other fields have generally discussed this phenomenon, suggesting that much of human activity is tacit and inarticulate, though nonetheless meaningful and purposive (e.g., Taylor, 1985). However, observers have also suggested that tacit activity can be explicaded, at least to some degree, and made the subject of critical examination. Such critical examination can enable practitioners to become more aware of subtle but important aspects of their practices and provide an opportunity for the refinement of their skills. In this sense, a greater awareness of the importance of informal learning can provide designers with opportunities for professional development—in terms of higher quality work in the present and an increased fund of experience for the future.

Our observation has some resonances with Schön’s (1983) notion of reflection-in-action, as well as views of critical thinking that call for greater awareness of assumptions and practices (e.g., Osguthorpe & Osguthorpe, 2007). Following this line of analysis, instructional designers would seek more visible and more refined learning habits as they pursue enhanced professional skill. Greater attention to informal learning, in this sense, would entail designers’ self-reflections regarding a finished product, what they learned through the process, what factors hindered or facilitated their informal learning efforts, what they would have done differently, and so on, all for the sake of continuous professional development. Moreover, designers with this disposition toward their work would be more cognizant of informal learning in the midst of design work itself, allowing them to adjust their efforts to better meet the demands of complex instructional problems and gain important experience in the process. From this perspective it might be hypothesized that the most capable instructional designers are, at least in part, the most capable informal learners; and novice designers who are the most willing to learn, in the sense we have described here, may become the most highly-capable designers.

References


A Review of Emotion Regulation in Intelligent Tutoring Systems

Mehdi Malekzadeh1*, Mumtaz Begum Mustafa1 and Adel Lahsasna2

1Multimodal Interaction Research Lab, Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia / 2Artificial Intelligence Department, Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia // malekzadeh@siswa.um.edu.my // mumtaz@um.edu.my // lahsasna@hotmail.com

*Corresponding author

(Submitted September 11, 2014; Revised January 26, 2015; Accepted March 18, 2015)

ABSTRACT

Having improved emotional (affective) state may have several benefits on learners, such as promoting higher cognitive flexibility and opens the learner to discovery of new ideas and possibilities. On other side, negative emotional states like boredom and frustration have been linked with less use of self-regulation and cognitive strategies for learning as well as increases in disengaged and disturbing behavior during learning. In the area of computerised learning, several researchers strongly agree that intelligent tutoring systems (ITSs) would significantly improve its performance if it can adapt to the affective state (emotional state) of the learners. This idea has spawned an important trend in the development of ITSs, which are systems with the ability to regulate a learner’s adverse emotions. In the present study, we discuss the existing studies that have implemented different emotion regulation strategies such as coping strategies and implementation of these strategies in the domain of intelligent tutoring system (ITS). The results of the review show that applying emotion regulation strategies during computerised learning may produce more optimistic emotions as well as better learning gain.

Keywords

Emotion, Emotion regulation, Emotion regulation strategy, Emotion coping strategy, Learning, Intelligent tutoring system

Introduction

A common view of emotions is that they are generated as a results of human’s judgment about the world and initiated by individual’s appraisal in response to and interaction with stimulus, such as material that the individual is learning (Desmet, 2002; Lazarus, 1991). Recent findings in neuroscience and psychology found that emotions are widely related to cognition, influencing various behavioural and cognitive processes, such as attention, long-term memorizing, decision-making, and so on (Ahn & Picard, 2005). Researches on emotion and learning suggest that positive emotions (affects) have a vital influence on various cognitive processes relevant for learning, such as information processing, communication processing, decision-making processing, negotiation processing, category sorting tasks, and creative problem-solving processes (Erez & Isen, 2002). Positive emotions promote higher cognitive flexibility and allow the learner to discover new ideas and possibilities. In addition, as a function of positive emotion, cognitive processes may be more flexible that result in greater creativity and improved problem-solving (Isen et al., 1987). Emotion also influences memory, where positive emotional state improved recall and it served as effective retrieval cues for long-term memory in many experiments (Isen et al., 1978). Reciprocally, negative emotional states like boredom and frustration have been linked with less use of self-regulation and cognitive strategies for learning as well as increases in disengaged and disturbing behavior during learning in the class (Isen, 2001). Thus, emotions, governed by proper attention, self-regulation and motivational strategies result in positive effects on learning, and lead to better achievement among the learners (Pekrun, Goetz, Titz, & Perry, 2002).

In traditional learning environment, a teacher maintains a sympathetic relationship with learners to facilitate the development of positive emotions. For instance, students who feel happy generally perform better than students who feel sad, angry, or scared (Connor & Davidson, 2003). This relationship also exists in a computerized learning environments and researchers of computer science in education field had studied techniques of artificial intelligence to make the educational systems more customized to the emotional state (affective states) of students (Jaques & Vicari, 2007).

Intelligent tutoring system (ITS) is a computer-based educational system that provides individualised instructions similar to a human tutor. Typical ITSs determine how and what to teach a student based on the learner’s pedagogical state to enhance learning. As experienced human tutor manages the emotional states of a learner to
motivate him or her and to improve the learning process, researchers also have augment the learner model structure in ITSs to determine the emotional state of learners (Neji, Ben Ammar, Alimi, & Gouardères, 2008). Researchers endow ITSs with the ability to detect learners’ unpleasant emotional states (e.g., confusion, frustration, etc.), respond to these states, and generate appropriate tutoring strategies as well as emotional expressions by embodied pedagogical agents. These emotion-sensitive ITSs aspire to narrow the interaction bandwidth between computer tutors and human tutors with the hope that this will lead to an improved user experience and enhanced learning gains (Aghaei Pour, Hussain, AlZoubi, D’Mello, & Calvo, 2010; Klein, Moon, & Picard, 2002).

In embedding emotional state reasoning into ITSs and intelligent learning environments, there are two main issues that are faced by the developers. First is determining the emotional states of the target learners, and second is determining factors that causes those states as well as how to respond and regulate negative emotional state (Avramides & Du Boulay, 2009; Du Boulay, Rebollo Méndez, Luckin, & Martínez-Mirón, 2007). To deal with the first issue, researchers (e.g., Aghaei Pour et al., 2010) paid attention to the determination of students’ emotions. Despite the complexity associated with real-time emotion detection, several researches have embarked on learner’s emotion detection. However, not many researches that focused on the causes of favorable or adverse emotional state of learners and strategies for regulating them. If the ITS design or the feedback offered were not suited to individual user needs and character, the learner can be frustrated or bored. The challenge is therefore to help learners to regulate their emotional states so that positive states such as flow/engagement persevere, while negative states such as frustration and boredom are prevented or regulated (Zakharov, Mitrovic, & Johnston, 2008).

In this paper, we have reviewed the efforts taken in existing researches that are related to regulation of negative emotional states in users’ learning process using emotion-sensitive intelligent tutoring systems (EITs). We discuss researches that apply the emotion regulation methods in emotion-sensitive intelligent tutoring environment. The findings of this review reveal that utilization of emotion regulation strategies benefits the learners during the learning process.

Materials and methods

For purpose of this study, we have searched the electronic databases that are relevant to education, psychology, information technology and social science: (a) IEEE XPLORE, (b) ACM Digital Library, (c) Science Direct, (d) Springer Link, (e) ERIC (Educational Resources Information Center) and (f) Web of Science. Searches were restricted to peer-reviewed articles, written in English, and published between 2008 and 2014 (research over the last six years).

We only included original articles pertaining to empirical research that focuses on applying different techniques for managing the negative emotional state of user such as boredom, anxiety, and sadness to improve learning productivity of the learner during learning episode with computerized learning system. The exclusion criteria include articles published in languages other than English and the research studies that do not meet the inclusion criteria.

The framework of Emotion-Sensitive ITS (EITS)

The interest in Intelligent Tutoring Systems began in the late 1970s, where these systems employs effective intelligent algorithms that would optimally conform to the learner and formulate strategy that optimizes the learning. The late 1990s and the early 2000s witnessed an exciting infusion of ITSs that incorporated tutoring strategies such as error identification and correction, frontier learning (expanding on what the student already knows), student modeling (inferring what the student knows and using that information to guide tutoring), and natural language dialogs (Aleven & Koedinger, 2002; Anderson, Douglass, & Qin, 2005; Woolf et al., 2009). Around the same time affective computing was beginning to rise as a new and exciting research area. Affective computing is about creating technologies that can monitor and appropriately respond to the emotional states of the user (Picard, 2010). Affective computing is a sub area of Human-Computer Interaction (HCI), where the emotional states of a user (feelings, moods, emotions) are incorporated into the decision cycle of the interface to develop more influential, user-friendly and natural applications (Picard, 1999). Throughout the last decade, several ITSs (Woolf et al., 2009; Zakharov et al., 2008) have been developed to incorporate assessments of students’ cognitive and emotional states into its educational
and motivational strategies to manage student’s engagement, self-confidence, regulate negative emotions, and maximizes learning (Calvo & Mello, 2011).

An EITS is generally divided into two main components. The first component is automatic identification of a student’s emotional states. There are several methods of emotion recognition proposed in the literatures such as through facial expression, body gesture, speech, text, and physiological measurements. The learners’ emotions are modeled in the valence arousal space, which is a 2D model for emotion modeling. Arousal describes the physical activation, varying from low to high, while valence refer to pleasantness or hedonic value, varying from negative to positive. Emotion such as stress, for instance, is modeled as high arousal and low valence, while joy and elation would be high arousal and also high valence (Schlosberg, 1954). The emotion classifier is trained on the recorded data using various sensors as student interact with ITS. Recorded data is annotated by multiple human judges including student (self-judgments) and trained judges. Labelling is required for supervised learning systems. In multimodal emotion detection, the emotion recognition component uses a decision-level fusion algorithm where each channel (conversational cues, face, posture, etc.) independently provides its own recognition of the learner’s emotional state. These individual recognitions are infused with algorithm that selects a single emotional state and a confidence value of the detection.

The second module responds to a user’s actions by adapting a teaching strategy based on pedagogical state (e.g., knowledge level, learning speed) and emotional state of learner. In EITs, agents continuously track student cognition, behaviour, or emotion and offer students with support based on individual differences along these parameters (Baker et al., 2006; D’mello & Graesser, 2012; Rebolledo-Mendez, du Boulay, & Luckin, 2006; Woolf et al., 2009). The agent behaviour and responses can be considered as a type of formative feedback to learners, and agents often propose a diversity of formative feedback strategies (Rebolledo-Mendez et al., 2006). Generally, the feedback strategy for emotion management can be domain dependent (e.g., providing hints and definition related to the course content) and domain independent (e.g., providing empathy or encouragement or requests to stop undesired behaviour).

In EITs, agents simulate human tutors by synthesizing emotional elements through the generation of speech, facial expressions, and other gestures. Using animated pedagogical agents allow EITS to offer sophisticated, real-time problem-solving advice and active emotional support with solid visual appeal. In addition, agents can motivate learners to interact more regularly with agent-based EITS and consequently increases the quality of a learner’s training over periods of months and years (Lester et al., 1997; Zakharov et al., 2008). In EITS, agent responses to the student’s action based on the learner’s cognitive and emotional state using a set of rules. Each rule has a set of feedback messages determining the agent’s verbal response. In addition, each rule includes a numeric value which triggers a change in the agent’s emotional appearance. For instance, when the learner answer correctly, the agent

![Figure 1. The architecture of EITS](image-url)
responds with a joyful smile together with an admiring message (Zakharov et al., 2008). Figure 1, depicts the architecture of EITS.

**Emotion coping and regulation**

In psychology, the concepts of emotion coping and emotion regulation are addressed to manage user emotional states. The emotion coping expends conscious effort to solve personal and interpersonal problems such as stress and conflict, and seeks to master, minimise or tolerate them (Lazarus & Folkman, 1984). Based on the work of Gross (1998), emotion regulation concerns with the ability to reduce high levels of emotion arousal and the capacity to change user’s feelings. Emotion coping focuses on decreasing negative emotion experience, whereas emotion regulation addresses increasing and decreasing both positive and negative emotions (Gross, 1998). Therefore, emotion can be regulated by using emotion coping as well as emotion regulation strategies by focusing on reducing negative emotional experiences. Lazarus (1991) classified emotion coping strategy in two different categories:

- **Problem-focused coping strategy**: Solving of the problem that causes the emotional situation such as providing definitions and examples related to the course content to learner during learning.
- **Emotion-focused coping strategy**: Reduction and management of the intensity of negative emotions caused by a stressful situation such as tutor provides encouraging statements during learning.
- Gross (1998), divides emotion regulation strategies into two categories: antecedent-focused and response-focused. Antecedent-focused strategies (i.e., situation selection, situation modification, attentional deployment, and cognitive change) occur before an emotional response is fully generated to influence an emotional state. Response-focused strategies (i.e., response modulation) occur after an emotional response is fully generated. The following statements describe these strategies (Gross, 1998).
  - **Situation selection**: Avoids or creates an emotionally relevant situation.
  - **Situation modification**: Modifies a situation to change the emotional state.
  - **Attentional deployment**: Distracts one’s attention away from a state.
  - **Cognitive change (Reappraisal)**: Reinterprets the meaning of an event.
  - **Response modulation**: Attempts to directly influence experiential, behavioural, and physiological response systems.

**Review of related works**

We have reviewed prominent research studies in the area of ITSs based on the selection criteria stated earlier. It is worth noting that in most studies, researchers did not specifically state the strategies they have applied in designing the feedback component for managing user negative emotions. However, these strategies can be categorised as emotion regulation strategies and coping strategies.

D’mello and Graesser (2012) designed evaluated two systems; AutoTutor and Affective AutoTutor. AutoTutor is an ITS that helps students to learn complex technical content in Newtonian physics, computer literacy, and critical thinking. AutoTutor is quite effective in helping students learn by holding a conversation in natural language, simulating the pedagogical and motivational strategies of human tutors and modeling and responding to their cognitive states. The affect-sensitive versions of AutoTutor, called the Supportive and Shakeup tutors, are collectively referred to as Affective AutoTutor were also developed. The emotional sensitive version of AutoTutor is capable in detecting learner’s emotional states, regulating negative emotional states, and synthesize emotions of the animated pedagogical agent. The agent’s feedback has been designed based on reactions to the emotional states of boredom, frustration, and confusion. The agent’s action to students’ negative emotions were derived from two sources, which are theoretical foundation (attribution theory and cognitive disequilibrium during learning (Craig, Graesser, Sullins, & Gholson, 2004) and recommendation by pedagogical experts. The attribution theory addresses boredom and frustration using empathetic responses from the tutor. The cognitive disequilibrium theory is also applied to address confusion, when a learner enters a state of confusion. Staying in a state of cognitive disequilibrium for too long is not recommended and the tutor should display empathy to acknowledge the learner’s attempts and lead the learner out of the state of confusion.
D’mello and Graesser (2012) have experimented with 36 undergraduate students from a university in the U.S using three computer literacy applications. Proportional learning gains were computed with different type of AutoTutor (regular, Supportive, Shakeup). The ANOVA analysis did not show any significant influence of tutor type or tutor. However, there was a 0.18 sigma trend in favor of the Supportive tutor compared to the regular tutor and a 0.28 sigma trend for the Supportive tutor over the Shakeup tutor. The results also have shown that, Supportive AutoTutor was more effective than the regular tutor for low-domain knowledge students in some sessions and the students with more knowledge never benefited from Supportive AutoTutor. They advise that, system should not be supportive until the students need support.

In the Wayang intelligent tutor system proposed by Woolf et al. (2009), variety of heuristic policies to respond to a learner’s emotions (providing text messages, mirroring student actions) were used. They investigated five independent emotional variables, including frustration, motivation, self-confidence, boredom and fatigue. The tutor responded to these emotional states by providing empathetic responses, agent change voice and gesture, presenting graphs and hints, giving encouragement, attributing failure to external factors, and changing the scenario. These types of responses are considered as problem-focus coping strategy (providing graphs and hints) or emotional-focus coping strategy (empathy messages) and emotion regulation strategies like situation modification (change the scenario) and cognitive reappraisal (attribute failure to external factors). They have measured interventions in relation to their impact on a student’s affect, behaviour and learning.

Chaffar et al. (2009) recognised a learner’s emotional responses after some tutoring of data structure web courses. They simulated two situations for the users. For the first situation, the tutor used problem-focused actions (using examples or definitions to change the situation that causes the negative emotion) and emotion-focused actions (helping participants to change their way of sensing the situation) to alleviate the effects of any negative emotion produced. For the second situation, after providing evaluation marks to students, the tutor used three emotion-focused actions, including encouragement, recommendation and congratulation as a way to encourage students to improve their marks and their knowledge in the future. The results of the ANOVA test showed that learners need help in understanding instead of encouragement when they did not understand the course. Hence, using a problem-focused action during learning was proposed. The results revealed that recommendation and encouragement actions have positive effects on the emotional states of weak learners after receiving their marks.

Strain and D’Mello (2011), have analysed the effects of cognitive reappraisal (an emotion regulation strategy) on learners’ emotional states and comprehension scores during a reading comprehension task. First, they injected negative emotions to participants. Next, they manage their negative emotions using two forms of cognitive reappraisal (deep and shallow reappraisal conditions). Subsequently, in a web-based learning session, participants were asked to learn about the U.S. Constitution and Bill of Rights and then answer questions about what they had learned. The results show that the utilisation of cognitive reappraisal as an emotion regulation strategy lead to more positive activating emotions and better reading comprehension.

Zakharov et al. (2008), used agent in their ITS to respond to students’ actions. Agent’s response is managed by a set of rules made in relation to the students’ cognitive states and emotional states. Each rule determines the agent’s verbal responses as well as changes to the agent’s emotional appearances. For example, when a learner has submitted a wrong answer several times, the agent’s verbal responses include a list of errors, with the appropriate emotional facial expression. Making the student conscious of their negative states may distract them from their negative feelings and move them towards their goal. Zakharov et al. (2008) used emotion coping strategies and regulation in designing feedback to reduce the negative emotions of learners. To evaluate the effectiveness of using the emotional agent in EITS, they performed an experiment in an introductory database course, with the experimental group that uses the emotion-aware version of the agent, while the control group had the emotion-unaware version of the agent. Since the learning sessions with ITs were short, the researchers did not expect to observe significant difference in learning performance between experimental and control group. The comparison among different conditions was made based on the questionnaire responses. In general, the findings supported the presence of emotional educational agents, with the emotion aware agent having advantages over its non-emotional counterpart.

Mao and Li (2009), proposed “Alice” an IETS with an emotion agent tutor. Alice was capable of recognising emotional states of a learner through facial expression, speech and text, and could adapt to emotional states of the learner with synthesized facial expression (providing empathy), emotional speech synthesis and text produced by the Artificial Intelligence Markup Language (AIML) Retrieval Mechanism. They consulted human teachers’ on suitable
educational and emotional actions for each scenario that can be applied by the agent in different learning situations. Mao and Li (2009), believed that emotional-aware agents incorporated in ITSs can optimise the learner behaviour towards learners’ enjoyment of the learning situation, though they did not report the result of any type of evaluation on their proposed ITS system. However, Mao and Li (2010) have conducted a pilot study, where 100 students used their proposed system to investigate the critical factors that impact learners’ satisfaction when using EITs. It was found that the agent tutor’s pedagogical action and expressiveness of the emotion expression are two of the significant factors in learners’ satisfaction from EITs.

Tian et al. (2014) have proposed architecture of interactive text-oriented emotion compensation mechanism in e-Learning to compensate the lack of emotion interaction between teachers and students in e-Learning systems. Their framework, based on affective computing and active listening strategy, recognizes and regulates the e-learner’s emotions based on interactive Chinese texts. They analyse the textual interaction data such as chart-rooms for courses, online Q&A, and group discussion for emotion recognition. Tian et al. (2014) introduced emotion regulation based on active listening, which is a non-judgmental feedback to an emotionally distressed individual, focuses on providing feedback of the emotional content itself. Active listening is an effective approach to regulate one’s emotion in real life (McNaughton, Hamlin, McCarthy, Head-Reeves, & Schreiner, 2008; Nugent & Halvorson, 1995). Tian et al. (2014) have applied a text-oriented emotion classification method to identify the e-learner’s emotion after he/she types a sentence, which is the listening step in active listening strategy. Furthermore, the case-based reasoning algorithm is adopted to recommend a similar emotion regulation such as a text-based advice, when the e-learner’s emotion is in negative states such as boredom, frustration, and fury.

Tian et al. (2014) designed an Emotion Regulation Agent (ERA) which is the core of the whole emotion compensation mechanism in the proposed framework. It analyses the topics, speakers’ roles and the interaction features of user input texts for the purpose of emotion detection from the text and predicts each learner’s emotion trend. In addition, it decides on the preference of emotion regulation strategies, so that e-Learners’ negative emotions can be regulated. The emotion regulation strategy library was constructed by including successful emotion regulation case base, according to the classification of Gross’s emotion regulation strategies (Gross, 2001). The success cases and the corpus were collected and labelled manually. The successful emotion regulation cases and some typical cases offered by psychologists are standardized into structured case templates and stored in the case bases.

Finally, the computation of similarity between short sentences and emotion regulation instances is presented. As an example, in a scenario of student group discussion, three-learner writes in the chat-box that he/she is depressed because his/her teacher criticized his/her report. Topic detection and tracking method is used to identify whether this is a new event or a recipient’s input. If the sentence was identified as new event, the sentence will be labelled and the emotional state and event content will be extracted. Every successful emotion regulation instance is shown in a format such as Event_set, Event_type, Emotion_category, Regulation_strategy, in emotion strategy database. The former two elements of an EES (Event Emotion regulation-in-Stance) are formed as (“his teacher criticized his report,” “frustration”). They used this feature to match the former element of EES of instances in the emotion regulation strategy database (Tian et al., 2014). Based on the recommended successful instances, a listener who communicates with depressed-e-learner will suggest him/her or say something like “communicate with the teacher”. This will comfort/relax the e-learner’s negative emotions. They have not reported evaluation of a comprehensive experiment on the real learners using their proposed framework.

Rodrigo et al. (2012), studied the emotional states of students while using ITS for Scatterplots with and without an interactive software agent (embodied conversational agent (ECA). Scatterplot Tutor (Baker et al., 2006) is an ITS that teach learners how to make and interpret scatterplots. This tutor was originally designed as part of the Middle School Mathematics Tutor and previously found to lead more learning gains (Baker, Corbett, Koedinger, Evenson, 2005). Scooter the Tutor was added to a Cognitive Tutor for Scatterplots, Scooter (which is an agent) reacted to gaming behavior with a combination of metacognitive messages (including requests to stop gaming), expressions of positive and negative emotion, and supplementary exercises covering the material the student avoided through gaming. Scooter successfully lessened gaming behavior of learner and significantly influence on students’ learning as compared to the same tutoring system with no agent. Baker et al. (2005), repeated the same experiment with students from a high school in Quezon City, Philippines. The participants are between the ages of 12 to 14. Data were collected from 126 students (64 in experimental condition and 62 in control condition), using a quantitative field observation method originally proposed by Pekrun et al. (2010). The findings from Rodrigo et al. (2012), were not the same as in Baker et al. (2005). Although Scooter is well favored by students and improves student learning
outcomes relative to the original tutor, emotional states and transitions between emotional states (emotion dynamics) were very similar between students in both conditions. In other word, Scooter did not have a significant influence on students’ emotional states or their dynamics. Boredom, confusion, and engaged concentration persisted in both conditions almost equally. One of the main finding of their research which was confirmed by other researches in emotion dynamics (Baker, D’Mello, Rodrigo, & Graesser, 2010; Kort, Reilly, & Picard, 2001), is that student emotion within learning software is quite stable as a cycle, regardless of whether or not software agents are present; the same student often is experiencing the same affect in following observations (180 seconds apart). For this reason, they believe that the first educational intervention that fundamentally modify students’ moment-to-moment emotion during learning, especially avoiding vicious cycles and creating and fortify virtuous cycles, will have made a major contribution and a major difference to learners.

Table 1 shows the summary of comparison among these studies mainly based on emotion regulation strategies that were used as well as their results.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Application</th>
<th>Regulation strategy</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain &amp; D’Mello 2011</td>
<td>Web-based learning system</td>
<td>Applying cognitive reappraisal (deep and shallow reappraisal) as emotion regulation strategies</td>
<td>Leads to more positive emotions and better reading comprehension score.</td>
</tr>
<tr>
<td>Chaffar et al., 2009</td>
<td>A virtual tutor that teaches data structure</td>
<td>Problem-focused actions (providing an example or a definition) and emotion-focused actions (change its way of perceiving the situation)</td>
<td>Problem-focused action leads to induce positive emotion during the comprehension task. Recommendation and encouragement actions have positive effects on the learners' emotions after receiving their marks.</td>
</tr>
<tr>
<td>Woolf et al., 2009</td>
<td>Wayang Intelligent Tutor (teaching mathematic)</td>
<td>Emotion-focused coping strategies (providing empathetic responses, agent change voice and gesture and encouragement). Problem-focused coping strategies (present graphs &amp; hints) and Emotion regulation strategies (attribute failure to external, change the scenario)</td>
<td>The interventions are measured in relation to their impact on student emotion, behaviour and learning.</td>
</tr>
<tr>
<td>D’mello &amp; Graesser, 2012</td>
<td>AutoTutor</td>
<td>An animated pedagogical agent was used to regulate negative emotional states such as frustration and boredom based on Attribution theory, Cognitive disequilibrium and experts recommendation</td>
<td>Supportive (emotion sensitive) AutoTutor was more effective than the regular tutor for low-domain knowledge students in some sessions and the students with more knowledge have not benefited from Supportive AutoTutor. Learning gains for the Shakeup and regular tutors were almost the same.</td>
</tr>
<tr>
<td>Zakharov et al., 2008</td>
<td>Intelligent Tutor system (teaching database design skill)</td>
<td>Problem-focused coping strategies (presenting the list of learner errors) and Emotional-focused coping strategies (change in the agent’s emotional appearance to empathise with the learner)</td>
<td>Based on learner’s opinions, ITS equipped with emotion-aware version of the agent has advantage over its non-emotional counterpart. However, there is no expectation for observing significant difference in learning performance measures because of short learning session.</td>
</tr>
<tr>
<td>Authors</td>
<td>Type of System</td>
<td>Methodology</td>
<td>Conclusion</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mao &amp; Li, 2009</td>
<td>Intelligent e-learning system</td>
<td>Emotional-focused coping strategies (adapt to emotional state of learner with facial expression generation)</td>
<td>Emotional-aware agent in ITS influence the mood states of the learner, or create positive impression. Agent tutor’s pedagogical action and agent tutor’s expressiveness are two important factors in learner satisfaction from using EITSs.</td>
</tr>
<tr>
<td>Tian et al., 2014</td>
<td>Chinese Text-Based e-Learning system</td>
<td>Using emotion regulation based on active listening to provide sincere, non-judgmental emotional content feedback to an emotionally distressed individual</td>
<td>A depressed e-learner will receive some sincere feedback according to the recommended successful instances. This will comfort/relax the e-learner’s negative emotions. However, they have not reported evaluation of a comprehensive experiment on the real learners using their proposed framework.</td>
</tr>
<tr>
<td>Rodrigo et al., 2012</td>
<td>Intelligent tutoring system for Scatterplots</td>
<td>An interactive software agent (Scooter) reacted to gaming behaviour with a combination of Emotion-focused coping strategies (metacognitive messages including requests to stop gaming, expressions of positive and negative emotion) and problem focused strategy (supplementary exercises)</td>
<td>The students are attracted to the agent (Scooter) and it enhances student learning outcomes relative to the original tutor, however, the Scooter did not have a significant influence on students’ emotional states or their dynamics. Boredom, confusion, and engaged concentration persisted in both conditions (with and without embedded agent for emotion regulation) quite equally.</td>
</tr>
</tbody>
</table>

**Discussion, conclusion and future work**

Arousal of negative emotional states such as anxiety, frustration, and boredom in learning environments as well as computerised learning such as ITS is inevitable. This situation may be due to the mismatch between the learner’s character and needs as well as the available functionality in ITS, inadequate interface implementations, system limitations, lack of flexibility, occurrences of errors and crashes. These factors contribute to the learner’s emotional state and filtering out negative emotion is difficult. These negative emotions can have severe consequences on students’ metacognitive and cognitive processes and their learning gains. Thus, ITSs should be equipped with the ability of guiding learners to regulate their negative emotions to achieve positive learning outcomes. Hence, in this paper, we investigated the findings of eight prominent researches related to emotion-sensitive computerized learning systems by applying different emotion coping and regulation strategies on the feedback provided to learners. We were interested to identify the strategies that are successfully and effectively implemented into EITSs to regulate negative emotional states of learners. We also study potential emotion regulation strategies that can be effectively implemented in ITSs.

Based on our findings and learner’s opinions, ITS equipped with emotion-aware agent has advantage over its non-emotional counterpart. Emotional-aware agent in ITS may influence the mood of the learner, or create positive impression. Agent tutor’s pedagogical action and agent tutor’s expressiveness are two important factors in learner motivation and satisfaction from using EITSs. The students are attracted to the agent and it enhances student learning outcomes relative to the original tutor. However, there are a few cases where the agent (e.g., Scooter) did not have a significant influence on students’ emotional states. Researches also show that emotion sensitive ITS is usually more effective than the non-emotionally sensitive for low-domain knowledge students compared to high-knowledge students. In addition, providing feedback in EITS based on emotion coping strategies were influential in regulating negative emotion of learners.
In most of the existing studies, emotion coping strategies including problem-focused and emotion-focused strategies are widely and effectively used in regulating negative emotion of learners in EITS. However, emotion regulation strategies proposed by Gross (2001) are rarely used in ITSs. Among the reviewed studies, Strain and D'Mello (2011), looks at the possibility of applying emotion regulation strategies such as cognitive reappraisal in intelligent tutoring system. As appraisal plays a serious role in the generation and experience of emotion, reappraisal changes the emotion experienced by the learner (Gross & Thompson, 2007). Findings from their experiments have demonstrated instructing learners to reappraise negative emotional states as they arise to help learners avoid negative affect and achieve better learning outcomes. In other words, instructed reappraisal (IR) involves instructing individuals to think of a negative situation positively to make the emotional experience less negative. Since IR strategies could help learners to become more engaged and improve their learning gains, such strategies could be implemented and employed in ITSs. Therefore, ITSs that are capable to recognize quickly the learners shift toward negative emotional state, can prompt learners to involve in IR strategies that will help them manage these negative emotional states and become affectively engaged. The limitation of IR strategies is that these methods have not been experimented on challenging topics such as mathematics or physics. Hence, effectiveness of instructing learners to reappraise negative emotional states during learning of challenging topics is still not known conclusively.

Our future research is to explore other strategies of emotion regulation like response modulation (suppressing emotional responses to stimuli) or attentional deployment to determine if these strategies are effective in changing negative emotional states of learners and their learning outcomes. Then they have potential to be implemented in the computerised learning environments.

References


Learner-Centered Blogging: A Preliminary Investigation of EFL Student Writers’ Experience

Ming Huei Lin

Department of English, Tamkang University, Taiwan // johnlinminghuei@gmail.com

(Submitted September 30, 2014; Revised January 10, 2015; Accepted February 12, 2015)

ABSTRACT

This study aims to investigate the effectiveness of integrating a learner-centered blogging approach into the EFL writing classroom in Taiwan. For this purpose, a 16-week experiment was conducted, involving an intact class of 18 university-level Taiwanese EFL student writers. During the experiment, the participants first created their own blogs on Lang-8, a free online language blogging platform, and then were encouraged to post at least seven entries on any issues that they wished to blog about. To better assess the effects of such an approach, both quantitative and qualitative inquiries were employed. Two writing tests were used to examine the participants’ writing performance, a questionnaire investigated their learning attitudes, and in-depth interviews explored the student bloggers’ learning experience in detail. The results suggest that the integration of learner-centered blogging into EFL writing instruction helps students develop writing skills as well as motivation and self-efficacy. However, their subjective blogging experience was mixed: although they conveyed enthusiastic support for the approach and strong belief in its effectiveness, this enthusiasm did not translate into much actual blogging activity. The paper concludes by offering suggestions for remedying the problems identified, and opening up avenues for further research.

Keywords

Blog-assisted language learning, Classroom blogging, EFL writing, Phenomenological analysis

Introduction

For several years, educators and researchers have tried to incorporate Web 2.0 tools in their instructional approaches and empirical projects in order to promote students’ writing skills and motivation to write (e.g., Lin, Lin, & Hsu, 2011; Richardson, 2009; Wang, 2014). Among the best-received applications have been blogs, which have successfully gained a foothold in education (cf. Sun & Chang, 2012) and have been widely seen as an effective approach to teaching writing to EFL/ESL students in particular (e.g., Halic, Lee, Paulus, & Spence, 2010; Lin, Li, Hung, & Huang, 2014; Vurdien, 2013; Ward, 2004). Specifically, it has been suggested that the use of blogs could augment student interactions via written language, enhance their learning motivation towards writing, and ultimately improve their writing performance (Arslan & Şahin-Kızıl, 2010; Blau, Mor, & Neuthal, 2013; Halic et al., 2010; Nguyen, 2012; Trajtemberg & Yiakoumetti, 2011; Taki & Fardafshari, 2012; Vurdien, 2013). Additionally, blogs have been viewed positively by EFL/ESL students as a novel, interesting, and creative platform on which they can freely blog their thoughts (Chen, Liu, Shih, Wu, & Yuan, 2011; Ducate & Lomicka, 2008; Nguyen, 2012; Noytim, 2010; Pinkman, 2005; Trajtemberg & Yiakoumetti, 2011; Wu, 2008).

Given the multiple benefits, it is no surprise that blogging has been widely perceived as a promising approach to facilitating teaching and learning (Churchill, 2009; Dyrud, Worley, & Flatley, 2005). However, its strong justification seems to have been established without taking account of some recurring issues in blog studies, in particular the increased work-load on blog teachers (cf. Levy, 2009, cited in Lin, 2014) and the inactive blogging models among students (Chiao, 2006; Lin, 2012; Wu, 2008). So far as the present researcher knows, these aspects still await further attention before the blogging effects claimed in most prior studies can be uncritically embraced in practical pedagogy in ESL/EFL writing classrooms. To investigate these issues, the present study conducted a 16-week learner-centered blogging project, where EFL student bloggers were given total responsibility over their learning process while writing blog entries. The language instructor involved was exempted from any extra work, which in itself allowed a fairer assessment of the effects of blogging on student writers. The project was expected to shed light on the following research question: Does integrating learner-centered blogging into the EFL writing classroom have an effect on EFL students’ writing performance and learning attitudes? To address this question, the researcher first verified whether or not student bloggers’ writing performance improved after the experiment, and then whether the participants experienced enhanced learning attitudes (motivation and self-efficacy in writing), the same qualities
commonly mentioned in previous studies as positive effects of blogging in writing classrooms. Next, the participants’ blogging experience was qualitatively explored via interviews, and their blogging patterns were related to these accounts. The researcher believes that these multiple examinations yielded valuable insights on the efficacy of the blogging approach.

**Literature review**

Classroom blogging has been reported to bring about varied effects on language students, including improved learning interests and attitudes, student interactions, and writing skills (e.g., Arslan & Şahin-Kızıl, 2010; Chen et al., 2011; Halic et al., 2010; Nguyen, 2012; Trajtemberg & Yiakoumetti, 2011; Vurdien, 2013). The multifarious blogging effects are likely to be ascribed to the nature of blogs as an open platform inviting an increased readership (Ward, 2004) and cultivating a strong sense of ownership/authorship (Oladi, 2005, cited in Warschauer & Grimes, 2007). These impacts have been argued to serve as an impetus for student writers to be more thoughtful and careful when working on the content and structure of their postings (Arslan & Şahin-Kızıl, 2010; Godwin-Jones, 2003; Noytim, 2010; Ward, 2004). Indeed, empirical evidence reported by Chen and Brown (2012) suggests that increased awareness of audience/authorship significantly affects student bloggers’ writing performance. These manifold higher levels of incentive and the advantages generated by writing blogs in turn probably explain much of the reason for broadly recommending blogging as an effective approach to teaching/learning EFL/ESL writing skills.

However, although plausible enough in themselves, these putative blogging effects and their mutual reinforcement do not seem to have successfully encouraged ESL/EFL students to engage more frequently or more willingly in blogging-supported writing activities (cf. Lin, Groom, & Lin, 2013). Although many studies have reported significant improvements to writing skills and enthusiasm about the idea of incorporating blogs into future classes, low blogging frequency has been consistently observed across studies (e.g., Chiao, 2006; Lin, 2012; Wu, 2008). It is true that this disjunction has not been reported in most “successful” blogging studies (e.g., Arslan & Şahin-Kızıl, 2010; Trajtemberg & Yiakoumetti, 2011). However, as Lin, Groom, and Lin (2013) observe, revisiting the student blogs from some of those studies (e.g., Arslan & Şahin-Kızıl, 2010; Lin, 2012; Trajtemberg & Yiakoumetti, 2011) shows that none of the student bloggers in these studies continued to post entries on their own or their peers’ blogs after the class projects finished.

To examine this disjunction between students’ positive attitudes to blog use in ESL writing classrooms and their inactive blogging patterns both during and after blogging assignments, Lin, Groom, and Lin (2013) interviewed a group of ESL students about their experiences after writing instruction/activities conducted using blogs. They found that blogging “increased awareness of [their own] limited linguistic ability” among participants (elementary-to-intermediate-level English writers) (p. 134). This dismaying self-consciousness may, as these authors observed, have raised students’ concern about the effort and time they needed to compose an English blog entry, thus discouraging them from blogging regularly.

Another stark issue needing attention is that the effects of blog-supported courses on student writers seem to have cost their language teachers an increased workload (Levy, 2009, cited in Lin, 2014). Teachers implementing blogging activities have often had to spend extra time and effort on providing students with support both for managing electronic content and for technical and knowledge development (cf. Churchill, 2009; Hourigan & Murray, 2010; Lai & Chen, 2011). This seemingly inevitable phenomenon is found to have an off-putting effect, even discouraging certain educators from any blog use in their classrooms (Lai & Chen, 2011). This may further explain the observation of Lin, Lin, and Hsu (2011) and Lin (2014): with the empirical research evidence from both short- and long-term blogging projects (18 and 36 weeks respectively), the improvements from classroom blogging on ESL students’ writing were too small to be justified “when the significant effort and amount of time spent on the design and maintenance of the blogs” were considered (Lin, Lin, & Hsu, 2011, p. 150). To meet the various challenges observed in the blog studies, an amended blogging approach, as suggested in this study, is clearly needed, in the hope of producing a different perception of blogging effects.
Methods

Sample

The experimental site was based in the English Department at a university in northern Taiwan. The study was conducted on one writing group composed of undergraduate English minors taking the compulsory course “English Composition (I),” which aims to enable students to develop well-organized paragraphs across various genres of writing in English. The group comprised 18 students (16 female), from rather mixed academic backgrounds, specifically, from the following nine departments: Industrial Economics (3 students), Insurance (1), International Business (2), Business Administration (2), Public Administration (1), Information and Library Science (1), Chinese Literature (1), Japanese (5), and Spanish (2). They were mostly aged between 19 and 20. Before participating in this experiment, the participants had studied English as a foreign language for between 6 and 10 years, and their proficiency ranged mostly from level A2 (four students) and B1 (eight students) to B2 (five students) under the Common European Framework of Reference (CEFR) for languages; one student had reached Level C2. All of the participants signed a consent form before joining the 16-week project, in which period their instruction time consisted of two 50-minute lectures per week.

It should be noted that the sample falls below the generally recognized minimum sample size for groups involved in experimental studies (i.e., 30; cf. Dörnyei, 2007; Groom & Littlemore, 2011, both cited in Lin 2014). Nevertheless, as Lin (2014) has reasonably argued, writing classes tend to have low numbers; given this practical restriction, a group of 18 should be considered adequate.

The treatment

The lectures were held in a computer lab where each student was provided with a personal computer connected to the Internet. All of the participants reported that they had access to either PCs or laptops in their dorms or at home, so after-class blogging activities were felt to be feasible.

In the first class lecture after the inception of the project, the students were introduced to the free blog server used in this project: Lang-8 (http://lang-8.com; see Figure 1). Lang-8 functions like many general blogging platforms (e.g., Blogger, Pitas, TypePad), enabling users to publish their journal entries easily online. Users can make comments (see Figure 2 for an example). Lang-8 also automatically creates a useful statistical record of the number of bloggers reading a writer’s entries, the number of comments received, and the numbers of other basic activities involved.

However, unlike most blogs, Lang-8 has the primary aim of achieving language exchange in a wide range of languages, with a particular focus on writing skills. This goal is realized by means of its nearly 1,000,000 (as of December 2014) registered users of diverse native language backgrounds worldwide, who may teach and learn their various written languages to and from each other (Manlove, 2007). To support this goal, Lang-8 incorporates a convenient gadget—Tracker, an application similar to the Track Changes function in Microsoft Word, which helps Lang-8 users provide corrective feedback directly on one another’s entries (see Figure 3 for an example of Tracker). This exceptional functionality of Lang-8 exempted the instructor in this project from needing extra time/effort to supervise, mark, or comment on students’ journal entries, thus addressing previous researchers’ concerns over the increased workload for language teachers of blog-supported courses. The use of Lang-8 as a blogging platform for the student writers in this study therefore successfully realized the idea of a “learner-centered” blogging approach and allowed objective observation of the effects of blogging on student writers.

After being familiarized with the functionality of Lang-8, the participants were asked to create accounts of their own and to produce at least one journal entry every other week (excluding pre- and post-test weeks, i.e., at least seven entries in all). To encourage frequent and active blogging, the students were informed that they could gain extra marks in the course if they posted more blog entries than required. Students were also encouraged to comment on peer work or that of other users on the open platform. With the “learner-centered” blogging approach, their instructor had no obligation to give comments or feedback on his students’ writing on Lang-8.
Figure 1. A snapshot of Lang-8 homepage

Figure 2. A Lang-8 page with comments on an entry
The tests

A pre-test and a post-test were administered to examine students’ improvement after the experimental treatment. Each test consisted of a 45-minute in-class writing task and a subsequent questionnaire. In the writing tasks, the students were asked to take sides on a general issue in student life, whereas the questionnaire explored students’ learning motivation and self-efficacy.

The pre- and post-test writing samples were assigned two holistic marks by two raters (the researcher and an experienced ESL teacher). Their scores had reasonable inter-rater reliability (Pearson’s $r = .70$ at $p = .001$ for pre-test results, and Pearson’s $r = .68$ at $p = .002$ for the post-test scores), indicating their validity for further inferential data analysis, for which their scores were averaged. Additionally, in order to demonstrate the validity of students’ writing performance and create a more comprehensive writing assessment, basic linguistic data on the participants’ writing samples, specifically on word tokens and word types, were also measured. These two measures have not only been found in the past to have statistically positive, significant correlations to students’ writing performance (Li, 2000;
Lin, 2009a, 2012) but have also been used as effective indicators of EFL students’ writing skills (e.g., Fellner & Apple, 2006).

The questionnaire adopted in this study was a 14-item survey designed by Lin (2012) to examine student writers’ self-efficacy and learning motivation. It has strong reliability overall (Cronbach’s $\alpha = .91$) and very strong construct validity (total variance = 72%). It is composed of three underlying factors, each of which individually has reasonable reliability and construct validity, namely, student writers’ (1) self-efficacy in writing content and organization (Cronbach’s $\alpha = .92$; variance = 30.17%), (2) self-efficacy in language use (Cronbach’s $\alpha = .89$; variance = 21.81%), and (3) learning motivation towards writing (Cronbach’s $\alpha = .80$; variance = 20.02%).

The interview

The qualitative data collected for this study consisted of transcripts of in-depth interviews with five student volunteers, completed on two different days but at the same location (the teacher’s office). Each volunteer was interviewed for around 30 minutes, using a set of open-ended questions corresponding to the aim of the research, namely:
- How did you like your blogging experience in Lang-8?
- What were your learning attitudes like in general when learning writing using the blogs?
- How do you evaluate the effectiveness of the blogging activities on Lang-8?
- Will you keep blogging on Lang-8? Why or why not?

All students were interviewed in Mandarin Chinese to ensure that they could express themselves clearly, precisely, and explicitly; the interviews were digitally recorded, transcribed verbatim, and later translated into English in order to be excerpted for the discussion below.

Data analysis

The data collected from the writing assessments and the questionnaire went through paired-sample $t$-tests using SPSS 17.0, to discover whether or not the participants’ performance on both writing and learning attitudes improved after the treatment. A descriptive analysis of the word counts was created using WordSmith 5.0 to complement the results generated, using inferential analysis.

The qualitative approach used here was a phenomenological analysis (Giorgi & Giorgi, 2008; Moustakas, 1994). As Lin, Groom, and Lin (2013) and Lin and Lee (2015) note, this approach involves a stepwise methodology for the researcher to refine the observations of individual experiences; this later forms a basis for developing generalized accounts of the collective understanding of the participants’ experience. Finally, descriptions extracted from individual and collective statements were interwoven to create a composite description addressing the nature and implications of the participants’ experience as a whole (cf. Moustakas, 1994; McNamara, 2005, cited in Lin, Groom, & Lin, 2013).

Finally, students’ blogging patterns have been an important issue discussed in previous studies (e.g., Chiao, 2006; Lin, Groom, & Lin, 2013; Wu, 2008), and students’ engagement in classroom blogging activities also seems to suggest, to some extent, their true perceptions of blogging (cf. Lin, Groom, & Lin, 2013). Hence, the activity and frequency of the interviewees’ blogging were also recorded in this study to supplement the quantitative and qualitative results.

Results

Quantitative results: Students’ performance in writing, learning motivation, and self-efficacy

The total number of word tokens and types gathered from the post-test writing samples (2,747 tokens and 668 types) was more than that for the pre-test writing collection (2,583 tokens and 625 types). The differences may seem small; however, given that the word tokens and types were produced in “timed” writing tests, this slight improvement may
suggest that students become slightly more capable of expressing themselves (that is, they are able to use more words) after learning in an EFL writing course supported by a blogging project (cf. Fellner & Apple, 2006; Lin, 2009a).

An inferential paired-sample t-test consistently found a statistically significant difference between the pre- and post-test writing performance scores ($t = -2.29, p < .05$; see Table 1), suggesting that the participants experienced significant improvements after attending the blog-supported writing instruction.

In addition to this gain, Table 1 reveals an evident difference between the pre- and post-test questionnaire scores ($t = -3.12, p < .01$), indicating that, after the experimental treatment, the student bloggers also experienced enhanced motivation and self-efficacy. More specifically, the paired-sample t-tests for the three questionnaire subscales reveal that the participants’ self-efficacy for writing content and organization made the greatest contribution to this improvement, since a significant difference was not found for the two remaining factors ($t = -3.12, p < .01$; see Table 1).

### Table 1. Paired-sample t tests for the test results

<table>
<thead>
<tr>
<th>Items</th>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Degree of freedom</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing tasks</td>
<td>Pre-test</td>
<td>18</td>
<td>80.81</td>
<td>3.80</td>
<td>17</td>
<td>-2.29</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>18</td>
<td>82.67</td>
<td>4.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Questionnaire</td>
<td>Pre-test</td>
<td>18</td>
<td>46.33</td>
<td>4.94</td>
<td>17</td>
<td>-3.12</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>18</td>
<td>50.11</td>
<td>4.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy in writing content and organization</td>
<td>Pre-test</td>
<td>18</td>
<td>18.17</td>
<td>2.09</td>
<td>17</td>
<td>-3.12</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>18</td>
<td>20.39</td>
<td>2.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy in language use</td>
<td>Pre-test</td>
<td>18</td>
<td>14.00</td>
<td>1.94</td>
<td>17</td>
<td>-1.62</td>
<td>.123</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>18</td>
<td>15.00</td>
<td>1.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation towards writing</td>
<td>Pre-test</td>
<td>18</td>
<td>14.17</td>
<td>2.66</td>
<td>17</td>
<td>-.89</td>
<td>.384</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>18</td>
<td>14.72</td>
<td>1.81</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Qualitative results: Students’ blogging experience

For reasons of space, this section will present only basic data on interview length and blogging patterns for each interviewee and a main result: a synthesized description of students’ blogging experience as a whole.

#### Interview length, writing performance, and blogging patterns

Table 2 lists each participant’s interview, with its length; English proficiency level, writing performance during the project, and blogging patterns are also noted.

### Table 2. Overview of the interviews: Duration and writing/blogging performance

<table>
<thead>
<tr>
<th>Student/Gender</th>
<th>CEFR</th>
<th>Interview time (min: sec)</th>
<th>Scores in pre/posttests</th>
<th>Total entries</th>
<th>Views received</th>
<th>Correction received</th>
<th>Comments received</th>
<th>Correction made</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2/F</td>
<td>A2</td>
<td>22:14</td>
<td>77.5 / 78.5</td>
<td>7</td>
<td>62</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>S5/F</td>
<td>B2</td>
<td>28:52</td>
<td>86.5 / 91.0</td>
<td>7</td>
<td>147</td>
<td>7</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>S8/F</td>
<td>B1</td>
<td>24:04</td>
<td>80.5 / 87.0</td>
<td>7</td>
<td>69</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>S14/F</td>
<td>B2</td>
<td>31:35</td>
<td>83.0 / 81.5</td>
<td>5</td>
<td>38</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S18/F</td>
<td>B1</td>
<td>31:53</td>
<td>85.0 / 82.0</td>
<td>4</td>
<td>35</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>MEAN</td>
<td>A2-B2</td>
<td>27:43</td>
<td>82.5 / 83.9</td>
<td>6</td>
<td>70.2</td>
<td>5.4</td>
<td>4.8</td>
<td>2.4</td>
</tr>
</tbody>
</table>

The interviews averaged 27 minutes and 43 seconds, which was enough time to gather adequate information for qualitative data analysis. English proficiency levels for the interviewees mostly spanned CEFR B1 and B2. Also, as shown in the table, most interviewees made progress in their writing from pre-test to post-test (the exceptions were S14 and S18). In addition, more than half of them reached the required number of journal entries (7). On average
each interviewee received 70.2 views for their journal entries altogether (11.7 views per entry), 5.4 corrections per person (nearly 1 correction per entry), and 4.8 comments per person (0.8 comments per entry). Most interviewees showed no interest in correcting others’ entries, and no interviewee produced additional journal entries beyond the required minimum. Overall blogging frequency was considered moderate or better—given the lack of official supervision from their teacher.

The composite account of the student writers’ blogging experience

The interviewees consistently described their blogging experience as “fun,” “cool,” “interesting,” and “fresh.” They saw Lang-8 as “a convenient, helpful, and innovative platform” where they could freely express their thoughts and receive feedback from native English speakers. As S18 specifically commented, “the website (Lang-8) is cool! There are foreigners reading your stuff (blog entries) up there, and they correct them and make them more native-like!” Although the teacher of this project was mostly absent during the students’ blogging experiences, the great majority of the interviewees acknowledged that they picked up some useful vocabulary and grammar from other bloggers’ corrections and comments, and approved of Lang-8’s effectiveness for learning English through blogging.

Interestingly, however, this overwhelming positive evaluation failed to bring about much actual blogging activity; this seemed to be hampered by the amount of time, effort, and (advanced) language skills required to complete a blog entry. Specific remarks on this point included: “My schedule is too full!” “I have no extra time for writing more.” “Time is a big problem!” and “It takes time to search for the right words I want to express myself.” S2 explained why she stopped writing after reaching the required number of entries as follows: “because [a blog entry] needs writing in English, which demands much time.” S14 similarly expressed concern about her language skills: “I gave up that draft (of a post on a particular event she had intended to write about) in English, and then wrote in Chinese and posted it on Facebook instead, because I was worried that my (English) vocabulary was not sufficient to precisely express my feelings.” Concerns about time and linguistic skills were also reflected in interviewees’ choices of topic: S18 chose easy stories or changed stories from her real life slightly in order to employ vocabulary that she had available. Similarly, S2 would reject a particular topic she wanted to write about if she realized that it was “too difficult to address in English.”

Interviewees’ concerns about their limited writing abilities and the effort needed to compose an entry were found to be associated with their pride, an association that further discouraged them from blogging. Specifically, S14 emphasized how she would make sure to the best of her ability that her entries were grammatically correct before publishing them, lest she lose face internationally online. S18’s strategy of choosing easier topics to write about was also a way to prevent herself from revealing the weaknesses in her English writing skills online: “My grammar is not always correct, and I don’t have confidence in [my] grammar, either.” This strong concern about losing face as a result of exposing their weaknesses may in turn explain why S14 and S18 failed to post the required number of entries. Even S2, who uploaded all seven entries, was not always happy at receiving many corrections, because of anxiety that “too many corrections from others make me look stupid.” Other interviewees did not specifically address these issues, but a comparable mentality can still be observed in the blogging strategies they chose. For example, S8 stopped reading others bloggers’ English entries after realizing that she would probably “pick up wrong writing skills in others' entries”, as she was “unable to tell whether others’ English writing is right or wrong.”

Interviewees’ overall positive assessment of the learning functionality of blogging may also be slightly attenuated by the fact that they sometimes had no clue why certain grammar points in their entries were corrected. S18 said: “it is cool to receive corrections from others … [but] sometimes I don’t understand why they made those corrections. I know some of my writing was probably wrong, but I don’t understand why it was wrong … Why can’t I put it that way?” S8 similarly noted that “[other bloggers] only revise my mistakes without providing explanations or comments.” S2 and S14 noted that most corrections they received were concerned only with grammatical usages and to a lesser degree word choice. Clearly, such help did not seem sufficient to some of the interviewees, in particular S8, who added that although the help with language use was useful, it was still insufficient for learning composition because “writing also requires structures and other skills.”

Finally, despite some critical remarks by learners themselves about the efficacy of blogging, every interviewed student blogger expressed great willingness to undergo more of this type of experience in future. Nevertheless, tracking their continued blogging patterns after this project, the researcher noticed that none of those interviewed...
students or any other student bloggers in this project tried to use Lang-8 or blogging of any kind. Only one male student uploaded an entry—8 months after the project ended.

In-depth discussion of the quantitative results and of students’ qualitative blogging experience is provided below.

Discussion

In this paper, the researcher employed a blended approach to examine the effects of integrating a learner-centered blogging approach into an EFL writing classroom where a group of Taiwanese undergraduates taking an English minor was involved. The results were mixed. Quantitatively, the participants’ English writing skills and overall learning attitudes significantly improved. The improvement in writing performance echoes the findings of previous studies (Chen et al., 2011; Fellner & Apple, 2006; Pinkman, 2005; Sun, 2010; Trajtemberg & Yiakoumetti, 2011; Tuan, 2010; Vurdien, 2013; Lin et al., 2014), where blogs are believed to supply a platform for enhancing students’ writing abilities. This finding in turn may support the student bloggers’ belief cited in Ward (2004), Fellner and Apple (2006), and Trajtemberg and Yiakoumetti (2011): that blogging could help them write better. The findings of enhanced learning motivation and self-efficacy toward writing are also in line with previous findings, from Taki and Fardafshari (2012) and İnceçay and Genç (2014): whereas the former authors found that student bloggers developed greater motivation towards writing, the latter reported that their participants experienced enhanced writing self-efficacy after blogging.

Qualitatively, as reported in many previous studies (cf. Fellner & Apple, 2006; Noytim, 2010; Trajtemberg & Yiakoumetti, 2011; Ward, 2004), the interviewees in this project also described some positive feelings and perceptions vis-à-vis blogging. These favorable attitudes produced an overall endorsement from the learners of learning English on a language blogging platform such as Lang-8. Along with the participants’ regular (if not active) blogging exercises, these conjoint strengths may in turn sustain the finding in the quantitative inquiry of blogging’s confirmative effect on students’ writing skills, motivation, and confidence. Given that the research method transferred most of the language teacher’s workload to other bloggers on Lang-8, the effects of using the learner-centered blogging approach in support of the EFL writing classroom seem to be fairly substantiated.

Although the efficacy of blogging seems robustly demonstrated, some of the bloggers’ concerns need further discussion here. First, the wider audience for their writings in the blogosphere did provide some helpful advice on grammar, but their awareness of this audience and increased self-consciousness also intensified some student bloggers’ anxiety about losing face due to their poor writing skills. On the one hand, this perception among the students is similar to the embarrassment that the student writers in the study of Lin, Groom, and Lin (2013) experienced when they blogged. On the other, however, this finding ran directly contrary to that of Lin et al. (2014), in which student writers claimed to experience less apprehensiveness when blogging online. Furthermore, the findings of this study also contradicted the generally acknowledged principle that an increased audience increases the positive effect on learning attitudes (cf. Chen & Brown, 2012; Oladi, 2005). To some extent, the present findings also conflicted with the common belief that e-journals, blogs, or computer-aided environments are a safer environment for self-expression than a regular face-to-face setting (Amir, Ismail, & Hussin, 2011; Myers, 2001; Parkyn, 1999; Peng & Hsu, 2008).

One possible reason for these diverse results is the different nature of the platform used by this study from that of most prior investigations. While most blog servers or similar online platforms are first and foremost used directly for communicative purposes, Lang-8 has an additional aim: helping its users with their linguistic skills through the embedded Tracker function. Devaluing to some degree its developers’ good intentions, this unusual emphasis served to expose student writers’ weakness in grammar to a larger readership. The negative perceptions that resulted could have influenced student bloggers’ confidence in their grammar skills, trapping them in a dilemma: to blog more or not to blog more (cf. Lin, Groom, & Lin 2013). This in turn may explain the difference here between the effects of the treatment on the two types of self-efficacy beliefs: that is, significant gain was found in self-efficacy only for writing content and organization and not in self-efficacy for language use. While the former could have been cultivated through the free expression of the writer’s thoughts, development of the latter was probably held back by enhanced anxiety about openly displaying linguistic weakness.
In addition, it seems necessary to consider the suitability of the blogging approach as employed in this project for EFL student writers whose English proficiency is relatively limited. Although the interviewees in this investigation seemed to approve of the grammatical and other corrections they received, most of them could not understand why their grammar was corrected. They expected more detailed explanations of grammar rules or other aspects of writing and written English, but such feedback was rarely offered. As a result, they may have misinterpreted the reasons for making certain corrections and then practiced these misapprehensions in their further blogging or other English writing. Although the participants’ short-term improvement in writing skills may at present seem to override such a possibility, it would be helpful to conduct longitudinal investigations to verify this.

Furthermore, although an online communicative setting is believed to involve no time pressure on students and no interruptions from either teacher or classmates (Roed, 2003), the participants in this project all experienced great time pressure in relation to their blog posts. To some extent, they also experienced stress about possibly losing face from posting linguistically incorrect sentences to an online audience. Such stress interrupts and impedes their writing; they confessed how they tended to reconsider and change what they had first intended to write. One obvious reason may be that the participants in this project were English minors, whose weekly timetables were filled with many other courses to consider. This workload could have increased their anxiety about the time needed for blogging, in particular when they were tempted to take advantage of the attractive features of Lang-8 but were unable to gratify such longing fully.

Conclusion

Because blogging has become increasingly popular in the EFL writing classroom, many research studies have shown that this approach can bring a wide range of advantages to the task of helping students learn to write. The present researcher also found results supporting the effectiveness of learner-centered blogging on a platform with a specific focus on language learning. This finding is particularly significant not only in terms of its potential for improving EFL students’ writing skills and general learning attitudes, but also by its showing that language teachers wishing to run a blog-supported classroom for their students need no longer carry the extra workload that accompanies blog use.

Despite the seemingly optimistic contributions overall, some unanticipated responses from the student bloggers still need special attention when such an approach is considered or implemented. This includes students’ concern about losing face and worries about blogging a journal in English with their limited foreign language knowledge. The researcher must also acknowledge that differences of culture and educational system may also influence the results of this study, as compared to previous studies. Taiwanese students have long been used to a more didactic pedagogical approach, considering teachers the centers of authority; the interaction between teachers and students is mostly teacher-directed (cf. Lee, 2013; Lin, 2009b; Tamney & Chiang, 2002; Yang, 2009). This means that most Taiwanese students are probably used to learning passively (cf. Lee, 2011), for example, receiving well-explained answers from their teachers or taking account of teachers’ comments or corrections. Consequently, it may have seemed rather alien to some of the present participants to be taught or to learn in an open blogosphere, with feedback in the form of comments from people other than their teacher. This novel learning environment may have made the students uncertain about how to respond properly to feedback.

It should also be pointed out that the blogging approach may have varying degrees of (un-)suitability for students, depending on their learning styles. It would be interesting and useful for future studies to consider different learning styles when introducing tools with the specific functionality of Lang-8, for example. Given this study’s exclusive focus on the effects on students’ writing abilities of integrating a learner-centered blogging approach into EFL classes, it should also be noted that students’ ways of coping with or using comments was not discussed, not least for reasons of space. Future inquiries into these aspects would provide wider understanding of student bloggers’ experience.

Finally, because the learner-centered blogging approach was integrated here with in-class lectures, as in most previous blog studies, the effects of such blends—in particular, the quantitative inquiries—were also assessed as a whole. Future studies may contribute to this field by locating more exclusively the stand-alone effects on EFL student writers of learner-centered blogging, possibly by employing a control group. This would further verify and/or enhance the reliability of the study’s findings.
Acknowledgments

This project was sponsored in part by the Ministry of Sciences and Technology in Taiwan (NSC 103-2410-H-032-001).

References


Exploring Effects of Multi-Touch Tabletop on Collaborative Fraction Learning and the Relationship of Learning Behavior and Interaction with Learning Achievement

Wu-Yuin Hwang¹, Rustam Shadiev²*, Chi-Wei Tseng¹ and Yueh-Min Huang²

¹Graduate Institute of Network Learning Technology, National Central University, Taiwan // ²Department of Engineering Science, National Cheng Kung University, Taiwan // wyhwang1206@gmail.com // rustamsh@gmail.com // s8413896@gmail.com // huang@mail.ncku.edu.tw

*Corresponding author

(Submitted October 6, 2014; Revised December 25, 2014; Accepted March 29, 2015)

ABSTRACT

This study designed a learning system to facilitate elementary school students’ fraction learning. An experiment was carried out to investigate how the system, which runs on multi-touch tabletop versus tablet PC, affects fraction learning. Two groups, a control and experimental, were assigned. Control students have learned fraction by using tablet PCs while experimental students have learned fraction by using multi-touch tabletop. It was found that learning effect of multi-touch tabletop was greater than that of tablet PCs. Multi-touch tabletop enabled students to easily collaborate, to learn fractions as well as to practice their problem-solving skills. On the other hand, students tended to work individually when using tablet PCs; therefore, students’ collaboration was limited and their understanding of fraction concept could not be promoted. As multi-touch tabletop was found as more beneficial to students’ learning, this study has further investigated the relationship of research variables (i.e., learning behavior to use the system and interaction among peers) with learning achievement in multi-touch tabletop environment. According to results, the number of times that fraction cards were canceled positively correlated with the post-test results. The reason is that canceling cards enabled students to refer to symbolic, graphical and simplified representations of fractions; thus, students could analyze fraction-related problem more thoroughly and solve it more efficiently. This study revealed that “seeking help” and “giving help” are important variables in peer interaction, particularly, in tightly collaborative multi-touch tabletop learning environment and they could lead to better learning achievement. With obvious awareness of peers’ needs and availabilities in multi-touch tabletop learning environment, students have easily asked for help from and provided help to peers during problem-solving process. In contrast, control students have usually worked individually or in loose collaboration. Results of this study also showed that most students had positive perception toward the multi-touch tabletop system, high collaborative learning attitude and motivation. Based on above-mentioned findings, this study makes several implications along with conclusions and suggestions for the future research.

Keywords

Tablet PC, Multi-touch tabletop, Fraction, Learning behavior, Interaction, Representation, Perception

Introduction

Mathematics refers to acquiring the basic concepts of figures, shapes and quantity (Ministry of Education, 2008). It is suggested that mathematics is an important subject as it is extensively applied in many fields. Being able to calculate, organize and apply mathematics knowledge and skills in daily life becomes prerequisites in everyday life and workspace. Therefore, mathematics learning needs to be emphasized, promoted and reflected in the aims of national education’s curriculum. General guidelines of grade 1-9 curriculum of elementary and junior high school education (Ministry of Education, 2008) set the following goals and claimed that educators must guide learners to achieve them: (1) to develop take-away mathematics skill so that learners can easily apply learned knowledge in real life situations; (2) to cultivate a positive attitude towards mathematics; (3) to teach mathematics by taking into consideration needs and development levels of learners; and (4) to ensure that learners are able to appropriately utilize technology in mathematics learning and problem-solving processes. However, realization of such ideas may encounter some obstacles which can easily be found in mathematics classrooms. For instance, according to House (2006) and Yu and Chang (2009), in most Taiwanese classrooms, mathematics classes emphasize that learners memorize formulas and calculate well in the condition of not knowing why and how. Under such circumstances, learners seldom explore meanings behind mathematical concepts and they hardly develop multiple mathematical competencies. As a result, learners fail to understand mathematics concept completely and to solve related problems. Furthermore, this issue leads to learners’ negative perceptions toward necessity of mathematics skills and importance
of such skills in the real life. This issue is especially relevant when learners learn some difficult mathematics concepts, e.g., fraction (Booker, 1998). House (2006) and Yu and Chang (2009) suggested that to merely memorize information is not enough. Learners have to be taught in such way so that they understand mathematics concepts well enough, acquire facts and skills in a meaningful manner and have appropriate learning experience. In this case, when learners experience real-life mathematics problems outside of school, they will be able to solve them.

Research on mathematics teaching and learning suggests that representations (e.g., graphical and symbolical) are important mediums not only for communication and expression but also for learning mathematics concepts. According to related studies, some concepts can be easily remembered and comprehended if they are presented in different representations (Nakahara, 2008). If learners learn difficult mathematics concepts (e.g., fraction) and use appropriately variety of their representations, learners will understanding them much easier and deeper compared to a situation when the same concepts presented in a sole representation (Dreyfus & Eisenberg, 1996).

Another way to facilitate learning of difficult concepts is to learn them collaboratively (Hwang et al., 2014; Shadiev, Hwang, Chen, & Huang, 2014b; Shadiev, Hwang, & Huang, 2015). According to related studies, learning collaboratively enables learners to achieve more than they could do alone (Argyle, 1991; Wang, 2009). One reason to explain this is that learners take advantage of one another during collaboration in terms of their resources and skills; so learners may ask for help from peers or provide necessary assistance to others (Lehtinen et al., 2001). It is suggested that collaborative learning of fraction can be better enhanced with educational computing technologies, such as tablet PC (Chen, 2011; Kong, 2011; Looi & Chen, 2010) or touch-operated tabletop (Jackson, Brummel, Pollet, & Greer, 2013; Rick, Rogers, Haig, & Yuill, 2009; Schäfer et al., 2013). However, several issues still exist with respect to these studies. First, most previous studies have focused primarily on exploring the effects of one or another technology on fraction learning. That is, it is still unclear which of them, i.e., tablet PC or touch-operated tabletop, are more prominent and beneficial for fraction learning. Furthermore, not much attention was paid in related research on the relationship of learning achievement with various independent variables, such as system usage or interaction among peers. For example, what kind of learning behavior to use a technology correlates with learning achievement or which type of interaction among learners predicts learning achievement the most?

This study attempted to address these two main issues. First, a learning system, which runs on multi-touch tabletop as well as on tablet PC, was developed. Elementary school students were invited to use the system for fraction learning and practicing their problem-solving skills in collaboration with others. This study aimed to investigate how the system, which runs on multi-touch tabletop versus tablet PC, affects fraction learning. The relationship of research variables (i.e., learning behavior to use a system and interaction with peers) with learning achievement were also explored. Finally, this study analyzed students’ perceptions toward fraction learning with a system. The following research questions were addressed in this study.

- Do students who have used multi-touch tabletop during fraction learning perform better on the post-test than those who have used tablet PC?
- What is the relationship of research variables (i.e., learning behavior to use multi-touch tabletop and interaction with peers in multi-touch tabletop environment) with learning achievement?
- What are students’ perceptions toward using the fraction learning system in multi-touch tabletop environment?

Related literature

Fraction learning and mathematics representations

Fraction is an important concept in mathematics course of elementary school students (Ministry of Education, 2008). However, according to Booker (1998), fraction is not an easy concept to learn. If elementary school students cannot understand this concept, they will experience learning difficulties with other related mathematics topics in high school in the future.

Some researchers have pointed out that learners’ ability to simplify fraction or to use different representations of fraction is an important factor for mathematics learning and problem solving. Simplifying or reducing fractions means making fractions as simple as possible; this can be done by dividing numerator and denominator by the highest number that can divide both of them. Mathematical representation was defined as a way to capture an
abstract mathematical concept or relationship. According to Nakahara (2008), five most common types of mathematical representations are: (1) symbolic representation such as numbers, letters and symbols; (2) linguistic representation such as utterance used in everyday communication; (3) illustrative representation based on illustrations, figures and graphs; (4) manipulative representation such as teaching aids that work by adding the dynamic operation of objects that have been artificially fabricated or modeled; (5) realistic representation based on actual states and objects. Gagatsis and Shiakalli (2004) have suggested that different representations for the same mathematical concept can enhance learners’ understanding of that concept. That is, different representations enable learners to build a connection between representations and/or converse from one representation to another. Therefore, mathematics representations gained much attention in many related studies (Jackson et al., 2013; Kong, 2011; Looi & Chen, 2010; Rick et al., 2009; Schäfer et al., 2013). In related studies, fractions were represented with numbers and graphs and learners have solved fraction problems by using different representations.

Collaborative learning

Kirschner, Paas, and Kirschner (2009) defined collaboration as a process when learners learn together to pursue a common goal. According to Wang (2009), collaborative learning requires certain mutual and shared effort from learners, for instance, interaction. That is, learners need actively interact with each other in order to establish a common focus and achieve a goal (Shadiev et al., 2014a). Lehtinen et al. (2001) have suggested that learners take advantage of one another during collaboration in terms of their resources and skills; so learners may ask for help from peers or provide necessary assistance to others.

The role of collaboration in promoting learning has been well documented (Hwang et al., 2014; Shadiev et al., 2014a). For example, Argyle (1991) and Wang (2009) argued that collaborative learning enables learners to achieve more (i.e., extent and efficiency of learning can be increased) than they could do alone. Moreover, collaborative learning helps to develop learners’ social and communication skills and establishes social relationship among other learners and group cohesion (Johnson & Johnson, 1999). Most studies on collaborative learning emphasize awareness as one of its core dimension (Hwang et al., 2013; Shadiev et al., 2014a). Dourish and Bellotti (1992) defined awareness as an understanding of an individual of activities of others, which provides a context for his/her own activity. Antunes, Herskovic, Ochoa and Pino (2014) suggested that individual can be informed about specific aspects of group members through awareness, e.g., where group members are (workspace awareness), what they are doing (social awareness) and what they are interested in (situation awareness).

Computer supported collaborative learning

According to Dillenbourg (1999), computer-supported collaborative learning (CSCL) is a pedagogical approach to learning of two or more people who learn or attempt to learn something together via social interaction supported by modern educational technologies. During CSCL, learners are able to learn either online or in classroom by sharing and constructing knowledge together using technology as a common resource (Hwang et al., 2014; Shadiev et al., 2014b; Stahl, Koschmann, & Suthers, 2006). A considerable amount of research has been published on CSCL. Of particular interest to this study researches on CSCL in which tablet PCs or tabletop multi-touch technology were employed for learning fractions.

Applications of tablet PCs for learning fractions

Advantages of tablet PCs for CSCL were repeatedly reported in the literature. For example, Chen (2011), Guerrero, Ochoa, Pino and Collazos (2006), Huang, Huang and Wu (2014) and Looi and Chen (2010) suggested that with tablet PCs, learners can learn collaboratively by communicating with each other, sharing knowledge and helping each other to accomplish learning tasks. Besides, the instructor may monitor and analyze learners’ behavior and performance based on learners’ digital profiles. Tablets can potentially reduce time needed for the teacher to do tedious logistical work (e.g., grading quizzes or engaging learners in learning activities). Kong (2011) argued that a tablet PC can support, guide, and mediate cognitive processes of learners. Learners demonstrate higher levels of motivation for learning mathematics when they learn with tablets; they actively learn and interact with other learners.
as they are able to move around the classroom and to conveniently bring along their devices in order to share
knowledge and learning material with each other. Therefore, tablets have the potential to facilitate learners’
performance, particularly for challenging topics, and to enhance learners’ engagement and autonomy in the learning
process. Due to these advantages, tablet PCs were effectively employed in many previous studies.

Chen (2011) has explored how students learn arithmetic calculation, such as fractions addition and subtraction,
collaboratively by playing Arithmetic Puzzle game. One experimental group (i.e., students who played the digital
game) and one control group (i.e., students who played traditional game) participated in Chen’s study. Students in
both groups were divided into teams of five and asked to play the game. In the game, four arithmetic problems were
given to students. Control students were provided with fraction problems on paper sheets while experimental
students could solve the same problems by using tablet PCs. A feedback was provided to students after they solved
all problems: the control group received it from the instructor and the experimental group from the system. Results
demonstrated that the experimental group showed a significant improvement in learning fractions compared to the
control group. Chen (2011) argued that it was due to timely provided feedback. According to experimental students,
the feedback messages were instant and very useful for self-reflection, especially when problems were solved
incorrectly. Students mentioned that the feedback was important for learning as to improve their math problem
solving skills and to enhance knowledge about arithmetic concepts. Chen (2011) claimed that low-achieving students
accessed the feedback more frequently and as a result, low-achieving students made the most significant learning
progress compared to students of higher achievement.

Kong (2011) proposed the Graphical Partitioning Model (GPM) as a cognitive tool to scaffold students’ fractions
learning. The GPM features visual representation, graphical manipulation, and immediate feedback. These features
enable students to easily find equivalent fractions, convert improper fractions and mixed numbers, add/subtract
fractions and simplify fraction forms. One experiment was carried out to investigate the effect of the model on
student engagement and attainment. The experimental class has learned the target topic with the use of the GPM,
while the control class has learned under the traditional teaching approach. The results showed that the use of the
GPM enhanced student engagement in learning about common fractions, in terms of time-on-task. The results of the
post-attainment test indicated that experimental students performed better than control students. Kong (2011)
concluded that the GPM effectively supported students in gaining a better understanding of common fraction
concepts and mastering the procedures used to add/subtract fractions with “like denominators.” Furthermore, the
GPM effectively provided additional support to students in generating untaught knowledge of the
addition/subtraction of fractions with “unlike denominators.”

Looi and Chen (2010) have introduced a tablet PC Group Scribbles (GS) system to facilitate students learning
representational notations for fractions and ratios in mathematics class. GS provides (1) a private space in which
students can work individually by creating private scribbled posts and (2) a group space in which students can post
the work and position it relative to others’, view others’ work and take items back to the private board for further
elaboration. The essential feature of the GS is the combination of the private and group boards. The class of students
was divided into groups of four, and each group was engaged in a problem solving learning activity. Looi and Chen
(2010) analyzed group interaction process supported by GS, particularly, how knowledge is built and shared with
others. Results showed several evidences of uptakes across the representational space and time that helped members
of the group to develop their understanding. One example is the proximity of the posts to each other served a deictic
function as students could talk about and refer to something by posting near it. Another example is group members
manifested their understanding later in the sessions when they commented on other postings. Looi and Chen (2010)
concluded that collaboration within the group mediated by the technology led to collaborative knowledge
construction.

A tabletop touch-operated technology to aid mathematics learning

With respect to tabletop touch-operated technology, Jackson et al. (2013) suggested that it accommodates multiple
colocated users and enables them to simultaneously interact with digital objects on the surface. Tabletop technology
produces more equitable interaction of users with digital objects. According to Rick et al. (2009) and Schäfer et al.
(2013), multi-touch table enables colocated group members to collaborate more flexibly (to share, discuss and
reflect upon their own and each other’s ideas) than using single personal computers. Besides, compared to personal computers, tabletops are more likely to elicit contributions from all members of a group and encourage more equal decision making and problem-solving. Due to these advantages, multi-touch tabletop was employed in many previous studies to aid mathematics learning.

Jackson et al. (2013) designed fraction learning game on an interactive tabletop. Students were divided into groups of four to play the game and to solve mathematics problems. Students were positioned around the interactive tabletop with two students at the bottom of the screen and one student on each side. Each student’s respective position had a resource pool. This pool contained the various tiles that students could use to solve problems. In the middle of the screen, a problem was presented. Within each problem, there was a location to deposit tiles that students thought were the correct solution. Once students thought they had a correct solution, they would press the “Check Answer” button located above a problem. If their solution was correct, the game would then load the next problem. If their solution was incorrect, the game would clear the current solution. Jackson et al. (2013) examined the effect that the interactive tabletop has on elementary students’ mathematical achievement. Results showed that students’ mathematical performance increased. Jackson et al. (2013) argued that the interactive tabletop learning environment produces helping behaviors; those students, who were helping their peers to understand and to learn the material in efforts to solve the group problem, experienced a significant gain in math performance. Based on results, it was suggested that the interactive tabletop can be an effective instructional aid.

Rick et al. (2009) designed DigiTile construction kit for math learning (i.e., fraction) which runs on a multi-touch table. DigiTile provides with pieces of different colors and shapes. This enables two co-located learners, i.e., next to each other, to place pieces simultaneously using touch input on the central tile (a square grid of snaps) in the middle of a tabletop to create a colorful tile. A feedback mechanism on fraction tasks is also provided by DigiTile; it displays the fraction of the central tile corresponding to each piece. One experiment was carried out with elementary school students. Results showed that students who underwent the DigiTile session had significantly higher scores on the post-test compared to students who learned fractions in the traditional way. Rick et al. (2009) concluded that DigiTile aided collaborative learning of fraction. Co-located learners have worked together (e.g., one places pieces while the other watches the fraction representation), modeled behavior for their partners, and articulated strategies and concepts for each other. With almost no guidance from the researcher or teacher, students showed significant improvement in their understanding of fractions.

Schäfer et al. (2013) have developed a game-based multi-touch learning environment in order to support the understanding and practicing of core mathematical concepts. The game consisted of multiple learning and playing modes in which teams of students collaborated and competed against each other. For example, in the learning mode, animation gave students insight into a concept’s basic rules, students practiced with a concept and received instant feedback, quiz game widened students’ topical knowledge and formula enabled students to construct logical formulas. In the playing mode, students collaborated, cooperated or competed against each other in order to learn and practice concepts. Schäfer et al. (2013) carried out an evaluation of the environment. Results showed that the game environment was regarded as easy to use and highly helpful. Learners confirmed that the game approach increased their motivation to get into abstract and difficult learning materials. Learners also agreed that playing the game is highly motivating, and they fully enjoyed dealing with the subject in a game-based environment. They specifically stated that practicing their problem-solving skills with the game was much more fun than practicing with pen and paper.

The literature review of this study shows that most previous related studies focused on exploring the effects of tablet PC or touch-operated tabletop on fraction learning. However, it is still not clear which of these technologies is more prominent and beneficial for fraction learning, particularly in collaboration. Furthermore, not much attention was paid in previous research on investigating the relationship of various independent variables, such as technology usage or interaction among peers, with learning achievement. For example, what kind of learning behavior to use a technology correlates with learning achievement or which type of interaction among students predicts learning achievement the most? Therefore, this study attempted to address these two issues.
Method

Participants and experimental procedures

This study carried out an experiment to investigate how the system, which runs on multi-touch tabletop versus tablet PC, affects students’ fraction learning. Forty-eight fourth-grade elementary school students have participated in the experiment. A pre-test was conducted at the beginning of the experiment. Based on results of a pre-test, this study assigned students into two groups: twenty-four students in a control group and twenty-four students in an experimental group. Fraenkel and Wallen (2008) have suggested that experimental studies can be carried out with such number of participants in a group if they are tightly controlled; however, the number of participants should not be less than 15. Heterogeneous grouping was employed in this study to create a relatively even distribution of students with different learning abilities. The data related to four experimental students were excluded from the data analysis as they did not fully participate in the experiment. Therefore, the data of only twenty experimental students was analyzed. Both groups had the same amount of the instruction on “Equivalent fraction”: 20 minutes class, once a week, for three weeks. After each class, students were assigned one fraction problem. Each problem was of different difficulty level: beginner level (A) in week 1, intermediate level (B) in week 2, and advanced level (C) in week 3. Students in both groups were divided into small teams of three to work on problems. Control students solved problems by using tablet PCs while experimental students by using multi-touch tabletop. A post-test and interviews with all students were carried out in the end of the experiment. Besides, experimental students were invited for the questionnaire.

System design

This study employed a multi-touch resource exchange system (MRES) for the experiment (Hwang et al., 2013). The MRES system was installed both on tablet PCs and a tabletop multi-touch. This system was implemented as one application of Surface Application Framework (SAF) using Java language and it was run on Windows 7® with the Microsoft® operating system. A tabletop was 70 cm in height and a surface had the physical size of 140×110cm with the resolution of 640×512 pixels. Tablet PCs were Asus Transformer Pads.

![The multi-touch tabletop system interface](image)

*Figure 1. The multi-touch tabletop system interface*

The system interfaces installed on the tabletop multi-touch and tablet PC are shown in Figure 1 and Figure 2 respectively. Tabletop multi-touch system interface includes one central area and three personal panels (i.e., one for each of three users). Personal panels are located on the edges of tabletop, i.e., in front of users. Tablet PC system interface includes one central area and one personal panel. A personal panel includes cards with different fractions. A card includes symbolic and graphical representation of fraction. Cards are randomly generated by the system. Central area presents a problem that users need to solve. For example, it shows a fraction “3/6” and users need to push card/s with equal fraction/s from their personal panels to central area. Furthermore, “simplifying fraction” function was designed for users in the system (Figure 3). By using this function, students can simplify fractions. For example, 3/6
is equal to 1/2 (i.e., 3\(\div\)6\(\div\)3\(\div\)2\(\div\)3\(\div\)=1\(\div\)2) and 12\(\div\)24 is also equal to 1/2 (i.e., 12\(\div\)24\(\div\)=1\(\div\)2\(\div\)1\(\div\)12\(\div\)=1\(\div\)2), so through simplification, students can find out that these two fractions are equal. After a student has pushed a card from personal panel to central area, “Cancel” button appears on users’ personal panels in order to cancel this action (i.e., cancel fraction card). Simplifying fraction and cancel fraction card functions were the same for the system installed on tabletop multi-touch and tablet PC.

![Figure 2. The tablet PC system interface](image)

The online workspace of control and experimental students was not different. That is, all students worked on assigned problems in small teams facing each other; control students by using tablet PCs and experimental students by using tabletop multi-touch.

![Figure 3. Simplifying fraction and cancel fraction card on multi-touch tabletop](image)

**Research tools**

This study conducted a pre-test to evaluate students’ prior knowledge and a post-test to measure students’ achievement after learning period. The tests were designed by the experienced elementary school teacher based on learning material covered in this study. Item facility, item discrimination and distractor efficiency (Matlock-Hetzel, 1997) were considering by the teacher in the test items design. The pre-test and post-test included 15 items (Appendix 1 presents some sample items) and items of both tests had the same structure, equivalent difficulty degree but different content. Tests were scored on 100 point scale (with 100 as the highest score).

The system has recorded the data related to students’ usage of the system and the following research variables were derived: (a) total learning time, i.e., the time spent by a group on problem-solving process; (b) total time of fraction simplifying, i.e., the time spent by a group on simplifying fraction; (c) total cancel time of fraction card, i.e., the number of times when students canceled their cards; and (d) total error times: the number of times when students’ provided incorrect answers. The unit of each research variable on usage of the system was based on individual student.
A video camera has recorded control and experimental students’ interaction with each other. All videotaped records were coded, categorized and then analyzed following general recommendation of Shaer et al. (2011). The reason for adopting suggestions of Shaer et al. (2011) is that their work was similar to our study, wherein interaction of co-located collaborators working on multi-touch tabletop was investigated. Three experts were involved in this process and big differences in the coding and categorizing were resolved through experts’ discussions and a consensus settled. Inter-rater reliability of the coding was evaluated by using Cohen’s kappa. The analysis result exceeded 0.77, indicating its high reliability. Table 1 presents five research variables related to students’ interaction.

The questionnaire survey included two parts. The first part aimed to survey students’ perceptions toward the system. Following general recommendations of Chunga and Tan (2004), Davis (1989), and Davis, Bagozzi, and Warshaw (1992), four dimensions were covered in the first part of the questionnaire: perceived ease of the system use (PEU), perceived usefulness of the system (PU); behavioral intention (BI) to use the system, and collaborative learning attitude (CLA). PEU is the degree to which a user believes that using the system would be free of physical and mental effort. PU is the degree to which a user believes that using the system would enhance his or her performance. BI is hypothesized to be a major determinant of whether or not a user actually uses the system. CLA is a student’s positive or negative behavior towards collaborative learning by using the system. Similar questionnaire survey was employed in many other related studies to investigate students’ perceptions toward the system (Hwang et al., 2014; Hwang et al., 2013; Hwang, Shadiev, Kuo, & Chen, 2012; Shadiev et al., 2014a).

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Definition</th>
<th>Example from transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seeking help</td>
<td>A student’s request for assistance from peers.</td>
<td>Can you explain how to simplify fraction?</td>
</tr>
<tr>
<td>Giving help</td>
<td>A student provides an explanation about a problem or how to solve it.</td>
<td>You need to simplify two fractions to find if they are equal.</td>
</tr>
<tr>
<td>Verbal shadowing</td>
<td>Short responses to suggestions or assistance.</td>
<td>Thanks, I forgot to do that.</td>
</tr>
<tr>
<td>Strategy</td>
<td>Students suggest and/or discuss a strategy to solve a problem.</td>
<td>You work on the first fraction, I will handle the second fraction, and you focus on the last one.</td>
</tr>
<tr>
<td>Syntax</td>
<td>Utterances referring to collaborative process.</td>
<td>Do you work on this fraction?</td>
</tr>
</tbody>
</table>

The second part of the questionnaire focused on students’ learning motivation. Learning motivation is an inner process which can maintain a certain behavior to be continued. According to Gardner (2010), learning motivation is the most influential factor in learning. If students’ learning motivation could be aroused, they may listen more carefully in class and review lessons consciously after school to pursue better performance. Maintaining motivation for longer time may develop students’ learning interests (Keller, 2010). In this study, the questionnaire included four dimensions (Keller, 2010): Attention is aroused and sustained due to learning activities; Relevance of learning content to tasks; Confidence to complete learning tasks; Satisfaction about outcomes to an effort to complete learning tasks.

Twenty valid answer sheets to the questionnaire were obtained from 20 experimental students. This study has utilized a five-point Likert scale, anchoring by the end-points “strongly disagree” (1) and “strongly agree” (5), for students’ answers. Cronbach α to assess the internal consistency of the survey was adopted and the values exceeded 0.73 in all dimensions which demonstrated satisfied reliability of the items.

One-on-one semi-structured interviews contained open-ended questions in which control and experimental students were asked about the following: (1) their experience using the system during the experiment; and (2) their opinions about the impact of the system for collaborative tasks. Each interview lasted for approximately 30 minutes.

**Results and discussion**

First, this study investigated whether students who have used multi-touch tabletop during fraction learning perform better on the post-test than those who have used tablet PC. Second, the relationship of research variables (i.e., learning behavior to use a system and interaction with peers) with learning achievement was explored. Finally, this study analyzed students’ perceptions toward the system.
Analysis of learning effects

The means and standard deviations of scores of the control and experimental groups are shown in Table 2. This study investigated the difference between the pre-test and post-test of the control versus experimental groups by employed paired-samples t test. Results of the test showed no difference between the pre-test and post-test of the control group, $t = -0.048$, $p = .962$. However, there was a difference between the pre-test and post-test of the experimental group, $t = -2.436$, $p = .025$.

Table 2. Results of the pre-test and post-test and analysis of paired-samples t-test

<table>
<thead>
<tr>
<th>Groups</th>
<th>Pre-test</th>
<th></th>
<th>Post-test</th>
<th></th>
<th>Paired-samples t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>$t$</td>
</tr>
<tr>
<td>Control</td>
<td>72.37</td>
<td>15.99</td>
<td>72.50</td>
<td>15.96</td>
<td>-.048</td>
</tr>
<tr>
<td>Experimental</td>
<td>71.70</td>
<td>16.36</td>
<td>77.75</td>
<td>13.79</td>
<td>-2.436</td>
</tr>
</tbody>
</table>

This finding suggests that multi-touch tabletop was beneficial to students’ learning so that they could understand learning material and solve related problems better compared to students who learned by using tablet PCs. The following is an explanation of it. In both settings (tablet PCs and tabletop multi-touch), students have worked in small groups of thee; however, the way control and experimental students have completed tasks was different. Tablet PC is referred to as a personal device since it has a small screen and it is not designed to be shared with multiple students. Therefore, it was not easy to use tablets to solve problems collaboratively. So this is why students solved problems by using tablets individually and their collaboration was mostly “loose” (Tang et al., 2006; Shadiev, Hwang, Huang, & Yang, 2014a). According to Olson and Teasley (1996), during loosely coupled work, learners are only aware of others’ activity and decisions and their work proceeds in parallel without direct dependence on each other. Although Kong (2011), Looi and Chen (2010), and other researchers have suggested that tablet PCs have the potential to facilitate learning, slack learning behaviors of some students were also reported in these studies. For example, it was observed that some students seldom complete learning tasks with their neighboring classmates when tablets are presented in class; each student relies on his/her own individual work even though students discuss assignments as a group. Kong (2011) suggested that this is possibly because students are not actively encouraged to discuss with peers. On the other hand, Looi and Chen (2010) claimed that such pattern (i.e., to solve problems in solitude) was due to students’ seating arrangement.

Tabletop includes three personal panels and, just like control students worked on their tablets, each experimental student has used his/her own personal panel to solve fraction problem individually first and then in collaboration with others. However, in contrast to tablets, in tabletop all three personal panels were integrated into one shareable workspace. Furthermore, the size of tabletop was far bigger than that of tablet PC so that students could surround tabletop and share the common workspace. Under such condition, students had much better awareness of each other’s problem-solving progress, needs and resources. If one group member could not solve his/her problem, this will be easily noticed by others, and then, instant and in-time help from others would be provided. This enabled efficient collaboration; so students have solved problems not only individually but they have also discussed about problems and worked on them together. Tang et al. (2006) and Shadiev et al. (2014a) called such collaboration as tight. According to Olson and Teasley (1996), “during tightly coupled work, learners’ work is directly dependent on each other, and their work typically involves a number of interactions to complete the task.” Therefore, students who worked on a multi-touch tabletop collaborated more effectively and their understanding of learning material was better enhanced compared to students who worked with tablet PCs. This finding is in line with other related studies (Jackson et al., 2013; Rick et al., 2009; Schäfer et al., 2013). Previous studies suggested that multi-touch tabletop is more likely to elicit contributions from all members of a group and encourage more equal decision making and problem-solving compared to personal computers. Multi-touch tabletop allows co-located students to construct digital content together and this form of co-construction allows learners to share, discuss and reflect upon their own and each other’s ideas.

Abovementioned finding demonstrates that multi-touch tabletop is beneficial to students’ learning. Therefore, in the following sections, we have only focused on research variables related to the use of multi-touch tabletop, i.e., the system usage, students’ interaction in multi-touch tabletop environment and students’ perceptions toward the system.
Analysis of the relationship between research variables and the post-test

Pearson product-moment correlation coefficient was employed to measure significance of the relationship between research variables related to the use of multi-touch tabletop and the post-test. First, the relationship between variables related to the system usage and the post-test was analyzed. According to results, the fraction card cancel times for problems of level B and C had a positive correlation with the post-test, \( r = .455, p = .044 \) (level B) and \( r = .468, p = .037 \) (level C).

Students were interviewed to obtain an explanation for this finding. Students mentioned that those, who have canceled many fraction cards, tried to find correct solution more frequently. They have attempted to find solutions by reviewing fractions' representations in symbols and graphics as well as in simplified form. Students have examined fractions very thoroughly and also compared their multiple representations (symbolical, graphical, or simplified). Such behavior could help students to find correct answer as well as their mistakes and to extend students' understanding of fraction concept. This finding is in line with other related research (Dreyfus & Eisenberg, 1996; Gagatsis & Shiakalli, 2004; Nakahara, 2008). Previous studies suggested that a difficult mathematics concept, such as fraction, is learned and understood much easier and deeper when variety of its representations is used appropriately. Furthermore, students who cannot solve problems and cancel cards many times are likely to get help from others. Therefore, those students, who canceled cards more, learned concepts related to fractions better and obtained higher scores in the post-test.

Second, the relationship between variables related to students' interaction and the post-test scores was explored. According to results in Table 3, the pre-test scores significantly correlate with “Seeking help” \( r = -.438, p = .054 \) and “Giving help” \( r = .602, p = .005 \). This finding suggests that students with low prior knowledge are likely to ask for help from others while students with high prior knowledge are willing to help others. Furthermore, results showed that the post-test scores have a significant correlation only with “Giving help,” \( r = .576, p = .008 \). This finding suggests that students who were highly motivated to help others usually have better scores in the post-test, which is consistent with our above-reported findings. Finally, results showed that learning gain (i.e., the difference between the pre-test and post-test scores) had a significant correlation with “Seeking help,” \( r = .604, p = .005 \). This finding may suggest that students who asked for more help from their peers had better learning gain. In the interviews, students (mostly with low abilities) claimed that if they ask for help from peers and get it timely, they could clarify a concept immediately. Therefore, this study suggests that seeking help was helpful to learning gain, especially of low ability students.

| Table 3. Pearson correlation between interaction and results of the pre-test and post-test. |
|-----------------------------------------------|----------------|-----|-----|----------------|-----|-----|
|                                               | Pre-test       |     | Post-test      |     | Learning gain |
|                                               | \( r \)         | \( p \) | \( r \)         | \( p \) | \( r \)         | \( p \) |
| Seeking help                                  | -.438          | .054 | -.033          | .890 | .604           | .005 |
| Giving help                                   | .602           | .005 | .576           | .008 | -.171          | .471 |
| Verbal shadowing                              | -.318          | .171 | -.154          | .518 | .278           | .235 |
| Strategy                                      | .220           | .351 | .243           | .301 | -.022          | .926 |
| Syntax                                        | .334           | .150 | .427           | .061 | .038           | .875 |

The following is a reason to explain this finding. Students who have low pre-test results are more likely to seek help from others when engaged in problem-solving process. Students, who seek for help and get it, usually improve their performance and hence, they have high learning gain. If students have high prior knowledge and/or high learning achievement, they are inclined to help others. Such students become helpers. These students usually do not have high learning gain because their test scores (i.e., the pre-test and post-test) are high. Based on this finding, this study suggests that our approach (i.e., the application of multi-touch tabletop for collaborative fraction problem solving) is beneficial for students with low prior knowledge. That is, these students can get help from their peers and as a result they learning gain becomes higher. Our approach can eliminate a big gap between students with different prior knowledge and learning achievement. Multi-touch tabletop is the good tool for collaboration, particularly, when students who seek or give help are involved; required support can be provided by peers and found from others easily and timely. Furthermore, in tabletop environment, students can easily discuss shared fraction problems; although students have individual panels to work alone, they still need to decide together which fraction is correct or incorrect. This finding is similar to those obtained in other related studies with regard to benefits of collaboration in tabletop environment (Hwang et al., 2013; Jackson et al., 2013; Lehtinen et al., 2001; Shadiev et al., 2014a). For example, Rick et al. (2009) found that co-located learners have worked together, modeled behavior for their partners and
articulated strategies and concepts for each other. Lehtinen et al. (2001) have suggested that students took advantage of one another in terms of their resources and skills; so students asked for help from peers or provided necessary assistance to others. On the other hand, tablet PCs are personal devices with small screens so they could not be employed for collaboration as easily as tabletop technology. With tablet PCs, most students focus on individual tasks and students are limited to collaborate efficiently.

Third, a stepwise multiple regression analysis was conducted to predict post-test scores with research variables related to peer interaction. According to results, the post-test scores can be predicted by “Giving help” variable during problem solving with level C difficulty. This variable was significantly related to the post-test scores, \( F = 9.753, p < .006 \). The multiple correlation coefficient was 0.315, indicating approximately 31.5% of the variance of the post-test scores could be accounted for “Giving help” variable. Other variables did not enter into the equation. Thus, this study concludes that “Giving help” variable during tasks with level C difficulty degree effectively forecasts the post-test effect. The reason is that level C problems were more difficult. Students, particularly with low prior knowledge, were in a great demand of help when they have completed level C tasks. Students who had high prior knowledge have helped those students in need. Consequently, all students could benefit from giving help interaction. On one hand, low prior knowledge students could get assistance and complete tasks. On the other hand, students with high prior knowledge could review their solutions once again and improve them when assisting others.

This finding was reflected in other related studies (Hwang et al., 2013; Jackson et al., 2013; Lehtinen et al., 2001; Shadiev et al., 2014a). For example, Jackson et al. (2013) have suggested that students, who were helping their peers to understand and learn the material in efforts to solve the group problem, experienced a significant gain in math performance.

Students’ perception toward the system

The questionnaire survey analysis revealed that almost all items were ranked high in the dimension “Perceived ease of the system use” \((M = 3.96, SD = 1.21)\), “Perceived usefulness of the system” \((M = 4.27, SD = 1.05)\) and “Behavioral intention to use the system” \((M = 4.45, SD = 0.99)\). This indicates that students generally agreed that the system was easy to use, useful to learn fractions, and most students were highly motivated to use the system continuously after this study. Nevertheless, we found that a few students scored some items very low because they did not like to use computers for learning at all. Results demonstrate that almost all items in the dimension “Collaborative learning attitude” were ranked high \((M = 3.93, S D = 1.23)\). This suggests that the system could support students’ collaboration of learning tasks. According to results, most students scored high for the items related to their motivation, i.e., attention \((M = 4.38, SD = 1.01)\), relevance \((M = 4.22, SD = 1.06)\), confidence \((M = 4.25, SD = 1.12)\), and satisfaction \((M = 4.38, SD = 1.06)\). It suggests that, in general, students had high motivation to learn by using the system. Results of interviews’ data analysis also strengthened the finding of the questionnaire. The following content was derived from interviews.

- The fraction picture which system offered that could help me quickly know which the same fraction is.
- Through the fraction card conversion, I was able to do step by step, to understand the fraction process of expansion / reduce.
- The fraction picture is clearly, it can help me to resolve problems.
- I like to discuss with peers when I work on problem-solving; it’s interesting.
- When I discuss with peers, I can get some help from them.
- Discussion is very useful and it can enhance our understanding of the given problem.
- Through discussion, I can understand confusing questions and find solutions.

Similar findings were found in other related studies on multi-touch tabletop (Hwang et al., 2013; Jackson et al., 2013; Rick et al., 2009; Shadiev et al., 2014a). Students in those studies have also accepted the technology and demonstrated their willingness to use it for fraction problem solving.

Conclusion

Related studies have pointed out that multi-touch tabletop and tablet PC are beneficial for collaborative fraction learning. However, this study found that multi-touch tabletop brings greater learning effect compared to tablet PC.
Besides, this study found that the more students cancel their fraction cards, the better are their post-test results. Furthermore, “seeking help” and “giving help” were found as important variables during collaborative fraction learning in multi-touch tabletop environment. Finally, most students positively perceived the fraction learning system and they had high collaborative learning attitude and motivation.

Based on these results, this study suggests applying multi-touch tabletop for collaborative fraction learning. This technology creates better learning environment in terms of awareness of group members and tight collaboration. Students in multi-touch tabletop environment are able to ask for help and provide help to each other efficiently and effectively. Furthermore, this study suggests that symbolical and graphical representations of fraction should be extended by providing fraction simplification function. In this study, combination of these three features enabled students to solve problems better, find mistakes and enhance students’ understanding of fraction concept. Finally, based on abovementioned findings, this study suggests naming multi-touch tabletop as collaborative computer (CC), instead of personal computer (PC).

This study also offers several implications for educators and educational system developers. Related literature suggests that different representations can be employed to facilitate fraction learning. However, it is really vital for educators and educational system developers to know what a good collaborative environment for math learning is and how it can be created by employing multi-touch tabletop technology. First, the environment should enable users to be aware of peers’ needs and availability for collaborative math learning. Besides, the environment should make users to achieve tight collaboration through intensive interaction rather than loose or scattered interaction, such as in the tablet PC environment. Second, the environment should enable users to use fraction cards with symbolic and graphical representations collaboratively in order to solve math problems. When users use fraction cards to solve math problems, really vital and effective cognitive development processes or misconception corrections take place through peers’ discussion; therefore, “seeking help” and “giving help” in this study were found as important elements of students’ interaction to their learning. Finally, this study implies that the above-mentioned points should not be considered separately as they mutually assist each other. That is, both the peer awareness in multi-touch tabletop environment and usage of fraction cards facilitate students’ cognitive development processes or misconception corrections effectively and efficiently.

In the future, the scope of this study will be extended. That is, the experiment will be replicated or experimental sample will be increased and experimental procedure will be prolonged to make results more robust and consistent. This will also enable results to be used for broader generalization. Furthermore, the future study will extend our current application with integration of multi-touch tabletop and tablet PC for fraction learning. That is, students will be engaged in individual problem solving process by using tablet PCs first, and then, transform their learning portfolios on multi-touch tabletop display. In this case, individual problem solving process can be extended by students’ collaboration, such as sharing, discussing and reflecting upon their own and each other’s ideas.

Acknowledgments

This research is partially supported by the “International Research-Intensive Center of Excellence Program” of NTNU and the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301, MOST 103-2511-S-006-007-MY3, and MOST 103-2511-S-006-002-MY3.

References


Appendix 1

The pre-test

**Part A.** Please pair fractions of the same value.

![Fraction images]

**Part B.** Fill in blank.

1. There are 48 bars of chocolate on the table. If you take 6, what fraction of chocolate bars would remain on the table? (Please write equivalent fractions)
   Fraction 1 (             )  Fraction 2 (             )  Fraction 3 (             )
   \[
   \frac{4}{7} = \frac{( )}{21} = \frac{28}{( )}
   \]

**Part C. Multiple choice questions:** Please select correct answer.

1. (             ) Which of the following equations is correct?
   (a) \( \frac{4}{5} = \frac{12}{20} \)  (b) \( 1 = \frac{1}{10} \)  (c) \( \frac{3}{4} = \frac{10}{12} \)  (d) \( \frac{1}{2} = \frac{100}{100} \)

2. (             ) 20 puddings were prepared and you ate five. What is fraction of remaining puddings?
   (a) \( \frac{4}{10} \)  (b) \( \frac{6}{10} \)  (c) \( \frac{2}{5} \)  (d) \( \frac{15}{20} \)

**Part D.** Word problems (please write the calculation process, otherwise an answer will not be scored)

1. Andy took \( \frac{2}{6} \) of his saved money to buy books but later he put back \( \frac{9}{24} \) of his saved money. Did he take more or did he put back more?

2. Teacher bought some pencils, and give Jill \( \frac{2}{6} \), Ming \( \frac{1}{4} \), and Jianhua \( \frac{2}{3} \). Who got the most of pencils?
A Robot-Partner for Preschool Children Learning English Using Socio-Cognitive Conflict

Elvis Mazzoni* and Martina Benvenuti

Department of Psychology, University of Bologna, Italy // elvis.mazzoni@unibo.it // martina.benvenuti2@unibo.it

*Corresponding author

(Submitted October 17, 2014; Revised January 21, 2015; Accepted February 13, 2015)

ABSTRACT

This paper presents an exploratory study in which a humanoid robot (MecWilly) acted as a partner to preschool children, helping them to learn English words. In order to use the Socio-Cognitive Conflict paradigm to induce the knowledge acquisition process, we designed a playful activity in which children worked in pairs with another child or with the humanoid robot on a word-picture association task involving fruit and vegetables. The analysis of the two experimental conditions (child-child and child-robot) demonstrates the effectiveness of Socio-Cognitive Conflict in improving the children’s learning of English. Furthermore, the analysis of children's performances as reported in this study appears to highlight the potential use of humanoid robots in the acquisition of English by young children.

Keywords
Humanoid robot, Children, Socio-cognitive conflict, English learning

Introduction

This research project deals with a robot (MecWilly) designed to help preschool children (4-6 years) in their learning of a second language (English). The main innovation is the experimental setting chosen in order to obtain an improvement in the children’s English language skills. In previous studies of educational contexts, robots have typically been used as “teachers” to “structure” the learning process (Fridin, 2014). This also means that children viewed the robot-teacher as “other” (in the case of humanoid robots) or as an “artifact” like a computer or a book (in the case of robots with non-human features).

The link between Robotics and Psychology has presented a very interesting challenge for many scholars over the last fifty years (Kelley & Cassenti, 2011). Starting from the development of the first types of computer (the first Turing-complete machine, ENIAC, created by John Von Neumann in the middle of the 20th century), the simulation of human mental processes, and, later, of human decisions and actions, has lead robotics to become one of the most exciting fields for the evolution of human behavioral models. Throughout this process, many of the most important models and theories of Psychology have been examined, in particular Piaget's Constructivism and the related processes of assimilation and accommodation, Skinner's operant conditioning, the Vygotskian Zone of Proximal Development, Cognitivism’s Human-Information Processing (HIP) model, as well as the latest models of artificial intelligence, Connectionism and Neural Networks (Dautenhahn & Billard, 1999a; 1999b; Ziemke, 2001). Even though all these models have been looked at as part of the broad field of Human-Robot Interactions (HRI) or in Robot Development/Evolution, each model has also been analyzed in more specific and well-defined scopes of application. We can briefly summarize five areas of application for psychological models in robotics:

(a) The evolution of robot prototypes which simulate human mental and physical behavior;
(b) Robot-robot interaction in complex environments for simulating human social behavior;
(c) Human→robot and robot→robot interaction in which the human or one of the robots acts as the tutor/teacher/expert and the (other) robot is a learner;
(d) Human←robot interaction in which the robot acts as a mechanical “partner” which allows the human to try out (and/or train) certain abilities or social skills so that he/she may make progress;
(e) And finally a situation in which the human-robot interaction is characterized by a human who acts as a “tutor/teacher” for the robot and a robot acting as a counterpart in the learning process which, in the end, sees the convergence of improvement and knowledge acquisition influencing the human’s cognitive development.

The first two areas have been principally influenced by Skinner’s Operant Conditioning, Cognitivism, and more recently by Artificial Intelligence (AI), Connectionism and neural networks (including their recent evolutions). Behaviorism played an important role in the development of early robots, as this psychological model has simple
rules which can be easily fed into a machine. Thanks to these characteristics, Skinner’s teaching machines were very successful in learning environments and demonstrated the importance of feedback (reinforcement) and also of coherence and repeated reinforcement over time in maximizing the learning process. Applying behavioral rules to robotics (and informatics), i.e., determining the behavior of a machine on the basis of the effects of a particular type of behavior and its related reinforcement, helped to produce a set of “mechanical” machines that were able to react to their environment by simulating certain types of human behavior. Robots capable of more sophisticated actions (and quasi-decisions) have been developed using behavioral models derived from AI connected to Behaviorism, the so-called Behavior-Based Robotics (Arkin, 1998; Brooks, 1986a). Robots like Allen (Brooks, 1986b), Herbert (Connell, 1989) or Genghis (Brooks, 2002) were considered revolutionary with respect to the previous ones, although some authors (Sharkey & Ziemke, 2000) highlighted something of a return to the past (Behaviorism) rather than a move towards the bases of AI and Cognitivism. Even more highly evolved behavior can be found in robots based on models derived from Connectionism, neural networks, and evolutionary swarm robotics (Trianni, 2008), in which the latter represents a good example of studies devoted to the simulation of human social behavior by using robots interacting in complex systems or environments. Despite the sophistication and the evolution of the robots’ behavior, these first two fields of application of psychological models and theories to robotics are primarily focused on the evolution of the robots to replicate human mental functions and only secondarily to enhance human cognitive development and knowledge acquisition by means of robots, i.e., the focus of attention of the final two fields of application.

A field of application which represents a bridge between simulation and human-robot interaction is the third field of application (c) in which the interest is not only in the simulation of human behavior, but also in the interaction between a robot and a human to develop the robot’s behavior as a consequence. The psychological models involved in this field are those derived from the Social Learning Theory (Bandura, 1971), the adaptation processes of assimilation and accommodation (Piaget, 1985; Piaget & Inhelder, 1969), and signs mediation (Vygotskij, 1978). Experiments conducted by Bandura (1971) unequivocally demonstrate that children exposed to a video in which an adult performs certain actions with certain tools, when left alone with the same tools, are highly likely to perform the same actions. These results show two important processes influencing human learning:

- The importance of reinforcement and punishment in learning is not always directly connected to the human experience. There are three types of reinforcement: past (explained by Behaviorism), promised (those that are not carried out but could be should certain actions be performed), and vicarious (by looking at the consequences of an action performed by another human, we can learn what to do to receive positive reinforcement).
- Humans learn not only by direct experience but also by observational learning or modeling, i.e., by observing the positive or negative consequences of an action performed by another human considered to be significant.

At a computational level, and in defining a robot’s behavior, we can imagine that it is relatively simple to program imitative behavior in a robot. Thanks to sensors that detect inputs deriving from the environment and a computational algorithm which takes these inputs and transforms them into actions, a robot can perfectly imitate a wide variety of human behavior and emotions. For example, Saunders and colleagues (Saunders et al., 2007, p. 109; Saunders, Nehaniv & Dautenhahn, 2006), used a Pioneer P3-DX robot and a Kephera robot to investigate how a robotic control and teaching system using self-imitation can be constructed with reference to psychological models of motor control and ideas from social scaffolding seen in animals.

It is, however, more difficult to replicate the modeling process in a robot since it involves not a simple imitation, but also a prediction and decision about which type of behavior will have more positive consequences (or avoid negative ones) based on previous “observations” of others’ behavior (Nehaniv & Dautenhahn, 2007). The difficulties arise when we depart from behavioral logic and we consider the logic of actions and activities. From an activity theoretical point of view (Engeström, 1987; Nardi, 1996), we can separate behavior from activity (Nikiforov, 1990) and define the latter as being organized in three elements: activity, actions and operations (Leont’ev, 1978; Mazzoni & Gaffuri, 2009). Unlike activity, behavior is neither directed nor subordinate to a predefined aim or set of objectives, and it therefore does not involve the forms of reasoning which typically precede human activity. At the same time, human behavior is situational since it is effectively a reaction to a situation. Although an activity can at times depend on the situation, the activity can also control and restructure the situation in order to achieve an objective (Nikiforov, 1990). On the hand, since behavior has no objective, it is motivated only by situational factors, that is to say, stimuli provided by the environment. We can better understand this differentiation by following the reasoning of Leont’ev (1978) which suggests that human activity is motivated by an objective which can be realized by goal-directed
actions, which, in turn, are accomplished by operations which may not be conscious, and which respond to situational conditions (Mazzoni & Gaffuri, 2009).

Actions and cognitive development were the focus of attention of one of the most important psychological scientists, Jean Piaget. Piagetian Constructivism emphasizes the active (and adaptive) role of a child interacting with the environment, guided by previous mental schemas for interpreting the environment, and open to the assimilation of new types of behavior/knowledge and/or to the accommodation of those previously held. Piaget clearly recognized the importance of reflexes (and positive and negative reinforcement) in his concept of an action scheme (composed of three elements: recognition of the situation $\rightarrow$ specific action related to the interpretation of the situation, expectation of a positive result), though he did so without limiting it to an overly simplistic mechanism of stimulus-response (Ziemke, 2001). Piaget’s view has been, and is, very important for robotics since it proposes a number of schemas for interpreting the environment which guides interaction with that environment as opposed to a simple reaction to it. We can summarize Piaget’s idea using Von Glasersfeld’s Radical Constructivism (1995):

- Knowledge is not a passive process of reception by means of the senses or communication but it is an active process of construction played out by cognition.
- Cognition has an adaptive function which enables a human’s organization of the experiential world.

As underlined by Ziemke (2001, p. 164), this notion is, at least at a first glance, largely compatible with a lot of recent research into cognitive science, artificial intelligence (AI) and artificial life which is concerned with adaptive robots or autonomous agents and their formation of internal structures in the course of agent-environment interaction.

The fourth area of application refers to the same types of psychological models described above, but in this case knowledge acquisition (and cognitive development) is linked to the child, while the robot acts as a mechanical partner, sometimes serving as a teaching assistant (Fridin, 2014), and at other times proposing creative programming activities which “provoke” some types of enrichment (Bers et al., 2014). This field of application has been very important for the majority of studies interested in enhancing the social skills of children with autism. Here the behavioral principles of reinforcement, repetitiveness and coherence are particularly important as autistic children tend to avoid complex and uncontrollable situations typical of many social contexts (Werry & Dautenhahn, 1999). Having a partner who always reacts to our actions in the same way brings the situation under our control and permits us to be more confident in social interaction (Dautenhahn & Werry, 2000). At the same time, imitation and some types of behavioral modeling are important, as the robot begins by reacting consistently but, as soon as the child shows more confidence, the robot’s behavior, though remaining similar, begins to change (e.g., the robot mirrors the child’s behavior) (Robins et al., 2004; Dautenhahn & Billard, 2002). Furthermore, these studies are based on basic constructivist principles that allow children to interpret the situation and interact with the robot so that, in the end, there may be an improvement in their behavior (in particular regarding the confidence they have in their social skills).

So far we have been very careful to divide behavior from actions in order to underline, on the one hand, the importance of feedback and reinforcement for human learning and, on the other hand, the essential and active role played by action schemes in organizing human experiences for cognitive development and knowledge construction. This theoretical background makes sense if we begin with the assumption that human knowledge construction (and cognitive development) is based on interaction with the environment and, logically, with its tools. At the beginning of the 20th century, Lev Semenovich Vygotskij, the father of the present Cultural Historical Activity Theory (CHAT), contested the simplistic idea of human learning (and cognitive development) based on stimulus-response mechanisms, arguing that:

- human cognitive development begins in social interactions with others and only later is what we construct during social interaction interiorized in an individual’s cognition (e.g., thanks to Piagetian processes of assimilation and accommodation);
- the role of mediating artifacts, i.e., tools (e.g., a machine or a robot) and signs (e.g., language or action schemes), is fundamental as they allow humans to interact with each other or with their environment, interpreting it and constructing their experience and knowledge (Vygotskij, 1978).

One of Vygotskij’s (1978) fundamental concepts is that of the Zone of Proximal Development, i.e., the potential for improvement in individual performances determined by social interactions or, in other words, the difference between the results achieved by a child acting individually and those that he/she can achieve by interacting with an adult or
another, more expert, child (Mazzoni, 2014). This idea is the basis for the fifth type of application of psychological models to robotics and is also the main inspiration for our experiment. Here we focus on the social interaction (dialogue) between two or more humans and between humans and agents (like robots) as a precursor to cognitive development, as opposed to the interaction between a human and the context. As Dautenhahn and Billard (1999b) have demonstrated in their studies using the Robotic Doll “Robota,” focusing on social context and the dynamics of interaction can lead to interesting experiments which can contribute to socially intelligent robotic agents.

The basic idea of this fifth type of application is that children interact with robots through dialogue and from this dialogue they construct their knowledge. The robot reacts or provides feedback principally based on behavioral mechanisms, while children construct knowledge by means of Piagetian assimilation and accommodation processes. This “social” interaction, however, takes place in a controlled setting (made possible thanks to the fixed patterns of behavior programmed into the robot), and is seen as having the potential to enhance the children’s abilities and construction of knowledge. The main idea is that by promoting the “correct” type of dialogue (in terms of predefined interactive sequences), complete with adequate feedback from the robot, we may activate an enhancement in a child’s knowledge. From this point of view, the robot is not simply a “reagent,” but it must instead be designed with particular features and specific speech software in order to sustain socio-constructivist interaction with children and activate the learning and knowledge construction process. Since dialogue is essential, the implementation of adequate behavioral schemas in the robot also becomes a fundamental aspect in the activation of a Socio-Cognitive Conflict in the child (Butera & Darnon, 2010; Carugati & Gilly, 1993). From a Socio-Cognitive Conflict perspective, individual development is conceived as being the result of social interactions made possible by the simultaneous presence of different points of view, and by the consequent necessity to negotiate common meanings or objects (Mazzoni & Gaffuri, 2009).

Drawing on Vygotskian socio-constructivist ideas (Alimisis et al., 2007; Kim, 2001), this research project focuses on the use of the robot MecWilly as an outsider “friend” for playful interaction with whom children have to negotiate their ideas in what becomes a classic situation of Socio-Cognitive Conflict (Butera & Darnon, 2010). Previous experiences in the field of Robot Assisted Language Learning have already used robots (Han, 2012; Mubin, Shahid, & Bartneck, 2013) or humanoid robots as teaching assistant in language learning (Lee, Noh, Lee, Lee, Sagong, & Kim, 2011; Mubin, Shahid, & Bartneck, 2013; You, Shen, Chang, Liu & Chen, 2006), also to teach a second language in primary school (Chang, Lee, Chao, Wang & Chen, 2010). Even though this field of research is in its early stages (Lee et al., 2011) and few studies discuss the use of robots to facilitate the teaching of second languages (Chang et al., 2010), generally the results achieved are promising and suggest that robots could be effective teaching assistants to improve (Wang, Young & Jang, 2013) or learn a second language.

The aim of this research project is to demonstrate that having a technological artifact (robot) partner, can be just as effective as having a human (child) partner for a child who is learning English by acting within their Zone of Proximal Development. The experimental setting has children playing in pairs or with MecWilly to solve a task consisting of associating the fruits and vegetables with the correct English word. The two situations are constructed to induce a Socio-Cognitive Conflict in which children have to negotiate their ideas with the other child or with MecWilly to arrive at a shared solution. The basic idea is that this Socio-Cognitive Conflict will enhance the children’s knowledge of English from the pre-test to the post-test, as they will have been made to consider different points of view.

**Socio Cognitive Conflict**

We can define Socio-Cognitive Conflict (SCC) as an interaction in which individuals reorganize and restructure their respective points of view to advance in their cognitive development by means of discussing their ideas. Cognitive improvement depends on the negotiation of points of view in order to arrive at shared understandings and agreement (Butera & Darnon, 2010). In other words, the SCC is a process in which “dissent from one or several partners over a task in which learning is concerned may stimulate task-related cognitive activity and result in progress” (Butera, Darnon & Mugny, 2010, p. 36).

Synthetically, the procedure to induce a SCC is characterized by a problem-solving task (with children this is normally a ludic task) in which a couple of children are asked to discuss and negotiate their points of view to reach a shared solution to the proposed problem (Doise & Mugny, 1984). In the classic situation created by Doise & Mugny
(1984), the different points of view are caused by the two pupils faced with a plan representing a map in which pupils are asked to reproduce the same village constructed by the experimenter in another plan near the first one. Since the two plans are rotated by various degrees, and since the two pupils are seated in different positions around the plan (e.g., one in front the other), in order to correctly solve the task they have to understand that sitting in different positions means having a different perception of what is left, right, in front and behind. Only the coordination of these different points of view, by means of subsequent negotiations, allows pupils to approximate the correct solutions and, very importantly, to enhance their cognitive level. In this and further studies (Butera & Darnon, 2010; Doise & Mugny, 1984; Doise, Mugny & Perret-Clermont, 1975; Carugati & Gilly, 1993), many authors have shown the relevance of SCC in pupils’ knowledge enhancement in terms of cognitive thought. Particularly relevant is the fact that results underline that enhancement occurs in all participants in the negotiations and, crucially, independently from a positive model of reference (Doise & Mugny, 1984).

Material and methods

This research is an exploratory study aiming at evaluating whether a humanoid robot can, by means of the socio-cognitive paradigm, be as effective as a human counterpart in helping Italian children who have no previous knowledge of English, in learning English words. The main actor in the study is the humanoid robot, MecWilly (Figure 1), an ecological humanoid robot (constructed entirely from recycled materials) characterized by three principal features: the replication and recognition of certain human emotions, the ability to move in a number of different ways in interactions with children in non-fixed contexts, a complex integration of software and sensors for recognizing human language, objects, and environmental changes determined by human behavior.

The proposed study examines and compares two experimental conditions:
- a Socio-Cognitive Conflict between two children (child-SCC);
- a Socio-Cognitive Conflict between a child and MecWilly (robot-SCC).

The research hypothesis is that, in learning specific English words through SCC, having a humanoid robot as a partner is just as effective as the classic situation in which two children collaborate in order to reach a shared objective.

In order to achieve our aims, the children’s classroom (Figure 2) was divided into two areas: one for the experimental situation and the other in which the children who were not engaged in the test played with their educator. The desk used for the MecWilly-child experiment (the lower part of Figure 2) was near a hidden lumber room in which the computer to control the robot was placed.
Figure 2. The room during the test phase

Here, a research collaborator monitored the situation (by means of a webcam placed on MecWilly’s nose) and controlled the robot’s answers and interactions with the child during the experimental session. MecWilly’s software and sensors for recognizing human language require a noiseless environment. The test phase was unable to guarantee this characteristic as it took place in the children’s daily scholastic environment. Therefore we decided that a research collaborator would activate certain predefined and standardized answers (by means of push-buttons) after having heard the children’s answers and comments (by means of the microphone installed in MecWilly). In this way, the same predefined answers (language + facial and corporeal expressions) were used with all the children. To induce the SCC, MecWilly’s answers did not provide solutions to the problems, but they simply induced a doubt, such as in the phrase “ahh, your suggestion is interesting … but are we sure that it is correct? Could there be an alternative or do we think that this is the correct answer?” (MecWilly 1). If the child did not make any suggestions, MecWilly intervened with the following: “Hmm, it is not simple … do you have any ideas about which could be a possible match? Do you have any suggestions?” (MecWilly 2). When children suggested a match, MecWilly replied with the phrase mentioned above (MecWilly 1).

Participants

The class was composed of 13 children between the ages of 4 and 6, although only 10 children (6 girls and 4 boys) participated in the experiment, because the others were sick. The whole procedure was carried out over 3 days and took place in a kindergarten in Bellaria Igea Marina.

Experimental procedure

After having obtained parental consent for each child, the educator told the children about the experiment a few days before it began, introducing it as a game. The educator told the children a tale in which MecWilly would come to visit from England with a list of fruit and vegetables that some English children would like to receive from Italy. However, MecWilly does not speak English very well and his list is in English, so it will ask the Italian children to help him to match the pictures of fruit and vegetables with the correct English word in order to work out which items it needs to send to the English children.
For the test phases we used 20 black and white pictures of fruits and vegetables designed by a professional sketcher (so that the sketches were similar): banana, melon, carrot, pear, onion, tomato, strawberry, apple, salad, eggplant, orange, cucumber, zucchini, apricot, peach, pepper, watermelon, cherry, lemon, and grape. The whole experimental phase, without considering the introduction made by the educator, took place over 4 days (two days of pre-testing, one day for the test, and one for the post-test) and lasted 90 minutes per day (45 to prepare the desks and the materials, and 45 for the test phases). To maintain the idea of the game, and to keep the children busy while their classmates were doing the test, at the end of each task they had the opportunity to choose two of the items, color them, and then put them in a shared basket, so we could determine the favorite fruit or vegetable of Italian children.

Pre-Test

Before starting the real pre-test, we tested the children’s ability to recognize the 20 items which were placed on the two desks. The children were divided into two groups (one for each desk), each group having the same number of boys and girls. The research collaborator asked a child “what is this?” in Italian, by pointing to a specific item, and the child said the Italian name of the indicated fruit or vegetable. All children correctly recognized all the items.

After this first control test, the research collaborators (one per desk) called two children at a time, one to one desk and one to the other (Figure 2), and after reminding them of the purpose of the activity (i.e., to help MecWilly with the English names of the fruit and vegetables) they explained the task. Each child had in front of him/her the 20 items representing fruit and vegetables. The research collaborator would say the English name of a fruit or vegetable (none of the children knew any English), and the child had to choose the picture that in his opinion matched that name. The child had two chances at answering, the second only given if the child did not answer correctly the first time. However, the children were not given any feedback about whether they had answered correctly, and none of the items (not even those correctly recognized) were removed from the desk during the pre-test.

During the pre-test phase, each research collaborator made a note of whether the child was purposeful (active) in front of the task or reluctant (passive). As SCC is based on a negotiation of different points of view, we tried to balance the pairs of children in the test phase in order to control the possible acquiescence effect already described in previous literature (Mugny & Doise, 1978).

Test

The pre-test phase was very important as we used it to decide which 6 items would be used for the test phase. During the pre-test, we encountered a problem in the children’s inability to stay concentrated on the task despite it taking no longer than 15 minutes. We therefore reduced the items of the test phase from 20 to 12 because there is no limit on the number of attempts the children can make in this phase: to improve their knowledge of English it was important for them to be able to negotiate their points of view and, finally, by means of answering as many times as necessary, to find the correct answer. Furthermore, in order to stay within the 45 minutes initially defined, we selected the 6 items which would be in the test. The 6 items were selected based on the amount correct answers they received in the pre-test phase (Table 1). Items that were correctly identified between 5 and 7 times were defined “easy” (E), those which were never identified correctly, or identified correctly only once, were defined “difficult” (D), while items which were correctly identified between 2 and 4 times, were defined “medium.” We had to exclude items which were always correctly identified because they would not guarantee a SCC.

During the test phase two children were sitting at one desk (children-SCC → 6 children → three couples) and one child was sitting in the other desk with MecWilly (robot-SCC → 4 children). When describing the activity (the same as for the pre-test) to the children (and to MecWilly), the research collaborators, one per desk, highlighted that before giving the answer the participants (child-child or child-MecWilly) would need to collaborate and talk to each other in order to reach a shared answer. In the robot-SCC condition, before giving his suggestion, MecWilly would wait for the child’s answer then make some suggestions, but would never actually give the correct answer. Based on the results of previous studies showing that a correct model is not necessary (Doise & Mugny, 1984), the idea was that MecWilly would serve as a counterpart helping the child to reflect on his or her answer and solution to a problem. MecWilly therefore normally made comments in Italian such as “Mmmhhh, your answer is interesting, but are we sure that it is correct?! Is there something else that could be the correct answer or do you think that is the correct
The answer was always given by the child, by pointing to the picture on the desk corresponding to the name proposed by the research assistant. In the case of an incorrect answer, the “wrong” item was taken off the desk and then it was put back after the correct answer was given, so that the maximum number of attempts to achieve a correct answer was 12 for each item. Unlike in the pre-test phase, in the test phase the children always had to find the correct answer.

<table>
<thead>
<tr>
<th>Level of difficulty</th>
<th>Italian</th>
<th>English</th>
<th>Pre-test correct answers</th>
<th>Post-test correct answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy (E)</td>
<td>insalata</td>
<td>salad</td>
<td>5/10</td>
<td>9/10</td>
</tr>
<tr>
<td></td>
<td>limone</td>
<td>lemon</td>
<td>6/10</td>
<td>9/10</td>
</tr>
<tr>
<td>Medium (M)</td>
<td>pera</td>
<td>pear</td>
<td>3/10</td>
<td>8/10</td>
</tr>
<tr>
<td></td>
<td>arancia</td>
<td>orange</td>
<td>3/10</td>
<td>7/10</td>
</tr>
<tr>
<td>Difficult (D)</td>
<td>albicocca</td>
<td>apricot</td>
<td>0/10</td>
<td>5/10</td>
</tr>
<tr>
<td></td>
<td>pesca</td>
<td>peach</td>
<td>0/10</td>
<td>6/10</td>
</tr>
</tbody>
</table>

**Post test**

In the post-test phase we repeated exactly the same task as that of the pre-test phase, with specific attention paid to the 6 items used in the test phase for which we expected to see an improvement both in the word-picture association task and in the children’s knowledge acquisition of English words. In order to analyze their knowledge acquisition, when the word-picture association task was completed, the research collaborator asked the child in Italian to help him remember the English name of a fruit or vegetable (by pointing one of the 6 items used in the test phase): “After all these names of fruit and vegetables I’ve forgotten the English word for this ... do you remember it?” After having noted the child’s answer (correct or incorrect), the collaborator proceeded with the other 5 items from the test phase.

**Results**

As this is an exploratory study with a restricted sample, we use a non-parametric test (Wilcoxon) to compare the pre- and post-tests (see Table 2).

<table>
<thead>
<tr>
<th>Test</th>
<th>Easy</th>
<th>Medium</th>
<th>Difficult</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-2.333b</td>
<td>-1.890b</td>
<td>-2.251b</td>
<td>-2.699b</td>
</tr>
<tr>
<td>p</td>
<td>.020</td>
<td>.059</td>
<td>.024</td>
<td>.007</td>
</tr>
</tbody>
</table>

The results show the improvement in the word-picture association task, both for easy and difficult items, but this improvement is most clearly visible when we consider the total.

To compare the two conditions and analyze the effectiveness of a robot in Socio-Cognitive Conflict, we used the descriptive data for the different types of items and the mean of the pre- and post-tests (see Table 3).

The results show a clear improvement in almost all children and for all items, though the mean of the pre- and post-tests highlights a significant improvement in the robot-SCC and a less relevant improvement in the children-SCC condition.

Finally, in order to analyze the acquisition of English words, the number of items learned from the 3 categories of item (easy, medium and difficult) shows that most children improve their English knowledge and the three best performances from 3 children (10, 8, and 7) were in the robot-SCC condition.
Table 3. Pre-test and post-test results in the two experimental conditions

<table>
<thead>
<tr>
<th>Experimental condition</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Acquisition of English words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word-picture association task</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E M D Sum Mean E M D Sum Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Children-SCC</td>
<td>0 0 0 0</td>
<td>0 1 2 3</td>
<td>2+0+1=3</td>
</tr>
<tr>
<td>2 Children-SCC</td>
<td>1 0 0 1</td>
<td>2 0 2 4</td>
<td>0+0+0=0</td>
</tr>
<tr>
<td>3 Children-SCC</td>
<td>2 2 0 4</td>
<td>2 2 1 5</td>
<td>1+2+0=3</td>
</tr>
<tr>
<td>4 Children-SCC</td>
<td>1 0 0 1</td>
<td>2 0 0 2</td>
<td>2+0+0=2</td>
</tr>
<tr>
<td>5 Children-SCC</td>
<td>0 0 0 0</td>
<td>1 0 0 1</td>
<td>0+0+0=0</td>
</tr>
<tr>
<td>6 Children-SCC</td>
<td>2 2 0 4</td>
<td>2 2 0 4</td>
<td>1+1+0=2</td>
</tr>
<tr>
<td>7 Robot-SCC</td>
<td>1 1 0 2</td>
<td>2 2 1 5</td>
<td>2+1+1=4</td>
</tr>
<tr>
<td>8 Robot-SCC</td>
<td>2 1 0 3</td>
<td>2 1 1 4</td>
<td>2+2+1=5</td>
</tr>
<tr>
<td>9 Robot-SCC</td>
<td>0 0 0 0</td>
<td>1 1 0 2</td>
<td>0+0+0=0</td>
</tr>
<tr>
<td>10 Robot-SCC</td>
<td>0 0 0 0</td>
<td>2 2 2 6</td>
<td>2+2+2=6</td>
</tr>
</tbody>
</table>

Discussion

The obtained results show the effectiveness of Socio-Cognitive Conflict in improving the children’s performance both in the word-picture association task, and in the acquisition of English words. Although these results are coherent with the literature in the field of Socio-Cognitive Conflict, they also show that a humanoid robot can be as effective as a human counterpart in the knowledge acquisition process. More specifically, all the children but one in the robot-SCC demonstrated a post-test performance which was better than that of the children-SCC (both in the association task and the acquisition of English words). This result may be explained by considering the difficulty of controlling the relations and negotiations in the children-SCC setting, particularly with children in Piaget’s preoperational stage. The robot-SCC is more structured since the robot always answers in the same manner and thus makes children take the responsibility for giving an answer and reflecting on it. Furthermore, as another study (Kanda et al., 2004) making use of a robot to learn English has shown, one of the requisites for improving students’ knowledge of English is a relationship with the robot. In this study, the human appearance of the humanoid robot MecWilly, together with the animism typical of preoperational children (Opfer & Gelman, 2011), guaranteed a relationship between child and robot in the robot-SCC condition which was very similar to that observed in the children-SCC condition. From this point of view, we might ask whether only a humanoid robot is so effective or whether computer software (like in the study proposed by Huang, Liu, & Shiu, 2008), particularly a smartphone or tablet application, might also be able to reproduce the same improvement by means of Socio-Cognitive Conflict. A robot, particularly a humanoid robot, has the clear advantage of creating a situation which is very similar to that proposed in the classic children-SCC, while other types of technological artifacts would suggest a different type of relationship, more similar to that analyzed in classic studies of human-computer interaction.

Conclusions

We analyzed the effectiveness of a humanoid robot in improving children’s knowledge of English words and we concluded that it does seem to play a relevant role in this process, and to be more effective than a human counterpart. Despite these promising results, the sample is too restricted to suggest that they may apply to other contexts in which preoperational children are involved. Further studies with larger samples are needed to confirm these results, and to analyze whether other types of technological artifacts (such as smartphones or tablet applications) might be able to activate the same process of improvement. It would be particularly worthwhile to replicate the study worldwide in order to compare cultural similarities and/or differences between young children in their interaction with robots. In these future studies, an important robot’s feature that could be improved and tested is the automatic interactive dialogue with children through Socio-Cognitive Conflict paradigm.
Acknowledgements

We would like to thank very much Roberto Masini, the creator of MecWilly, to have accepted the proposal of this research and to have prepared and managed the robot during the experimentation. We would also thank very much the teachers of the kindergarten “San Giuseppe” in Bellaria Igea Marina, in which the study took place, particularly Gloria Giovannini who was actively involved in the project.

References


Language Learning in Virtual Reality Environments: Past, Present, and Future

Tsun-Ju Lin¹ and Yu-Ju Lan²*

¹Department of Applied Chinese, Kainan University, Taoyuan City, Taiwan // ²Department of Applied Chinese Language and Culture, National Taiwan Normal University, Taipei, Taiwan // tsunjulin@gmail.com // yujulan@gmail.com

*Corresponding author

(Submitted November 3, 2014; Revised May 4, 2015; Accepted May 7, 2015)

ABSTRACT
This study investigated the research trends in language learning in a virtual reality environment by conducting a content analysis of findings published in the literature from 2004 to 2013 in four top ranked computer-assisted language learning journals: Language Learning & Technology, CALICO Journal, Computer Assisted Language Learning, and ReCALL. Data from 29 articles were cross-analyzed in terms of research topics, technologies used, language learning settings, sample groups, and methodological approaches. It was found that the three most popular research topics for learners were interactive communication; behaviors, affections, and beliefs; and task-based instruction. However, the analysis results highlight the need for the inclusion of the impact of teacher. The data also revealed that more studies are utilizing triangulation of measurement processes to enable in-depth analysis. A trend of gathering data through informal learning procedures was also observed. This article concludes by highlighting particular fields related to VR in which further research is urgently needed.

Keywords
Virtual reality, MMOGs, Virtual worlds, Content analysis

Introduction

The recent rapid advances in and popularity of wireless communication and multimedia environments have resulted in language learning in a virtual reality (VR) environment [a so-called virtual learning environment (VLE)] receiving considerable attention in the past few years. In the early stage, VR generally involved only small representations of content area or domains designed to “bridge the gap between reality and abstract knowledge by the discovery method” (Lee, 1999, p. 72). In order to create such environments, the systems were commonly theme-based and comprised a full integration of artificial intelligence and a wide variety of social communication tools. Because of its features, VR was quickly introduced to promote authentic and immersive learning environments.

VR tools nowadays are a far cry from early two-dimensional (2D) text-based online VR environments, such as multiuser domains (MUDs) and MUD, object oriented (MOO), to which multiple users can be connected at the same time. Current VLEs have evolved to be more sophisticated and interactive with a high degree of visual appeal, allowing a wide range of interlocutors to communicate, cooperate, and compete through customized three-dimensional (3D) virtual spaces and avatars. Currently there are several VR tools in the market; Sykes et al. (2008) categorized VR based on its original design purposes into three types: open social virtualities, massively multiplayer online games (MMOGs), and synthetic immersive environments (SIEs). Open social virtualities, such as Second Life (SL), OpenSimulator, and Active Worlds (AW), allow the users to immerse themselves in a wide variety of social contexts, participating in individual and group activities, and letting them create and trade virtual properties and services with one another. MMOGs, such as World of Warcraft (WoW), emphasize more role-playing functions, and involve hundreds or even thousands of players cooperating and competing with each other simultaneously based on their selected race, class, or profession. However, Sykes et al. (2008) considered that the above two types of VR were originally developed for commercial and business use. In contrast, an SIE emphasizes its educational objectives by carefully incorporating pedagogy into the immersive spaces. Regardless of the original purposes for which VLEs were designed, researchers in the computer-assisted language learning (CALL) field have tried to employ pedagogical principles and practices that are innovative and theoretically grounded to understand the pedagogical values of VLEs in language learning.

Given the above background on how VR emerged and has been categorized, it is useful to define VR so as to reflect its development as a state-of-the-art technology. Based on a thorough review of the definitions proposed by other
researchers, Smart et al. (2007) offered the following core definition of VR: a system that aims to bring simulated real-life experiences, providing topography, movement, and physics that offer the illusion of being there.

**Review of VR and simulation research**

Several journal articles conducting meta-analysis have attempted to identify the characteristics and challenges of VR in technology-enhanced learning research. Lee (1999) analyzed 19 studies to determine the effectiveness of computer-simulated environments by investigating the relationship between 2 forms (i.e., pure and hybrid) and 2 modes of instructions (i.e., presentation and practice). That study revealed that within either the presentation or practice mode, hybrid simulation (which has expository instructional features) is much more effective than pure simulation (which does not integrate any specific directions and explanations). Schwienhorst (2002) examined the state of VR in second-language (L2) acquisition with the aim of identifying the common problems encountered and the role of VR in CALL research. Researchers in the 1990s seemed to agree that VR can promote constructivist activities and learner engagement (i.e., student involvement, self-monitoring, and self-evaluation). Additionally, VR can lead to “intrinsic motivation, more intercultural awareness, and a reduction of the affective filter” (Schwienhorst, 2002, p. 230). However, Schwienhorst (2002) claimed that VR research during that period was inadequate to allow any conclusive claims to be made.

Subsequent studies seemed to switch the focus to the effects of instruction involving computer simulations versus traditional instruction methods. Vogel et al. (2006) reviewed 32 studies and found that across all variables (e.g., age and gender) and situations (e.g., types of activities and the degree of image realism in the computer programs), the use of interactive simulations and games provided superior cognitive outcomes and more positive attitudes toward learning than did traditional teaching methods. Similar to previous studies, they also found that interactive simulations promoted self-directed learning and provided fail-safe learning environments. The authors also demonstrated that interactive simulations and games increase learners’ knowledge more than traditional teaching methods across all ages. Liao and Chen (2007) examined 29 studies performed in Taiwan and also found that instruction involving computer simulations had more positive effects on learning by students than did traditional instruction methods. Their study further suggested that 3D virtual presentation may be more effective than 2D for students’ learning due to the former providing more accurately simulated and authentic learning environments.

It should be noted that among the four above-mentioned studies, only Schwienhorst (2002) emphasized the use of virtual worlds for language learning. Moreover, although Schwienhorst (2002) claimed to have conducted a meta-analysis, the article was actually more of a historical overview of relevant research and aimed to shed light on the use of VR in L2 acquisition. This is understandable given that in 2002 there were too few language studies involving VR to allow a meta-analysis to be conducted (aimed at establishing statistical significance with studies on a particular topic that have produced conflicting results); moreover, the number of examined studies was not mentioned in his article.

The subsequent decade saw a focus on the research trends in technology-based learning. Hsu et al. (2012) conducted a content analysis on five SSCI journals published from 2000 to 2009, with the results revealing that the proportion of articles on VLEs, digital games, and learning enhanced through the use of “intelligent” toys increased significantly between 2000–2004 and 2005–2009, from 0.81% to 3.82%. However, VR is still reported to be one of the least published research topics in the technology-based learning field. Similar to Hsu et al. (2012), Wang and Vásquez (2012) also reported VR as being less frequently explored technologies compared to other Web 2.0 tools in L2 acquisition.

The above-described results indicate that studies that analyze the appropriateness, application, and practices of VR, and its influence on language education are urgently needed. This situation prompted the present study to systematically review and synthesize the literature on language learning in VR from 2004 to 2013 in order to determine whether this research field has produced any conclusive data during the past 10 years. Reviewing the trends in VR may help to identify research interests and gaps, and further provide reference data for future research directions. To identify the overall research trends, this study categorized previous studies into five categories: research topics, technologies used, language learning settings, sample groups, and methodological approaches. The following four research questions (RQs) were specifically addressed by this study:

**RQ1. What percentage of the articles published in the selected journals were related to VR?**
RQ2. What topics related to language learning in VR were investigated in these journals from 2004 to 2013?
RQ3. What technologies have been used in VR studies?
RQ4. What methodologies were applied in VR studies from 2004 to 2013? How did the methodologies change from 2004 to 2013?

Method

The studies to be included in the content analysis were identified by performing computer searches of journal databases. Since it would have been difficult to include all relevant journals in the analysis, the following four top CALL-specific and education technology-related journals were selected: *Language Learning & Technology (LLT)*, *CALICO Journal*, *CALL*, and *ReCALL*. Smith and Lafford (2009) evaluated these as the four highest quality English-language journals in the field of CALL. The examined journals were evaluated by 35 tenured CALL experts using a list of criteria, including the quality of the articles, significance of contribution to the field, review process, and originality. It was expected that a thorough analysis of high-quality articles in these 4 journals by the 35 experts would identify the most influential works in the CALL field.

In total, 811 articles appeared in the 4 journals from 2004 to 2013; this excluded columns, commentaries, book/media reviews, review studies, editorial materials, and letters. A content analysis was applied to these articles using the following keywords: simulation, VR, VLEs, social virtual worlds, MOOs, MUDs, and MMOGs. In a few cases where the keywords were insufficient, the researchers analyzed abstracts and full articles. This process yielded 29 empirical studies for inclusion in the analysis in this study. Based on the methodology of content analysis, descriptive statistics was utilized to classify the previously mentioned five categories, namely research topics, technologies used, language learning settings, research sample groups, and methodological approaches.

Research topics

Based on the methodology of inductive content analysis, the research topics related to L2 acquisition in CALL were first classified into four categories: learner differences, learning task, impact of the teacher, and environment. For each category, subcategories were classified. During the data analysis process, these categories and subcategories were refined continuously. The validity of these categories and subcategories was confirmed after all the articles were reviewed, as follows:

- **Learner differences**: Articles related to this research topic explore learner’s linguistic and sociolinguistic competencies. Some articles also discussed the concept of metalinguistic awareness in terms of how a learner integrates form, meaning, and function to keep oneself using the normal target language. This category was further divided into four subcategories: interactive communication, motivation, cultural awareness and intelligence, and language proficiency.

- **Learning task**: This category includes articles that discuss or examine different types of instructional strategies and approaches used in VR. This category was further divided into three categories: task-based instruction (TBI), collaborative learning, and problem-solving.

- **Impact of the teacher**: This category includes studies that outline teachers’ teaching experiences and perceptions in MUVEs. The two subcategories comprise task design, and perception and awareness.

- **Environment**: Articles that emphasize investigating the affordance of a VR environment were classified into this category.

Technologies used, language learning settings, research sample groups, and methodological approaches

First of all, the three types of VR proposed by Sykes et al. (2008) were extended to the following subcategories of research technologies used in this review: open social virtualities, MMOGs, SIEs, “others,” and “not specified.” Second, after conducting several rounds of discussions, the following six subcategories were identified for the language learning settings: first language (L1) learning, L2 learning, foreign language (FL) learning, language exchange, “others,” and “not specified.” Third, discussions among researchers reached the following consensus for the eight categories for the research sample groups: elementary school, junior/senior high school, higher education, teachers, adults, special needs, “others,” and “not specified.” It should be noted that the “adults” sample group
included all adult participants without specifying whether they were teachers or participants in higher education. Finally, the same researchers discussed the usage of the subcategories for the methodological approaches, which produced the following finalized subcategories: quantitative, qualitative, and a combination of quantitative and qualitative approach (quan.+qual.). The data collection for each article is described later.

Data coding and analysis

This study used the articles relevant to VR published in the four selected journals from 2004 to 2013 to investigate the research trends. Content analysis was utilized to classify the aforementioned categories.

A one-level coding process was used to categorize technologies used, language learning settings, sample groups, and research designs. Articles with only one major subcategory were coded to the designed subcategory, while those with more than one major subcategory or with subcategories other than the aforementioned ones were accounted as “others.”

As mentioned above, many of the analyzed articles addressed more than one research topic. Thus, the research topic was coded into two levels, with the primary and matched subcategories coded as the first and second research topics, respectively. Any matching subcategories after the second research topic were excluded, and the field of the second research topic was coded “none” for articles with only one research topic.

The coding process was undertaken manually by the researchers. To obtain more reliable outcomes from coding, three researchers (two professors and one research assistant) in educational technology helped to code these studies based on the aforementioned categories. These researchers each had from 5 to 12 years of language teaching experience, knew more than one language, and had taught online courses in more than two different countries previously. All the articles followed the same coding process. The results were compared using Pearson correlation coefficient measures of reliability, which yielded $r = 0.84$. Any discrepancies were resolved by discussion. After the initial coding process, a descriptive analysis was conducted to report the data.

Results

Percentage of articles related to VR

The analysis by year and journal (see Table 1) revealed that only 3.6% of the articles published in the four journals from 2004 to 2013 were related to the field of VR: 0.8% in LLT, 1.9% in CALICO Journal, 4.7% in CALL, and 6.3% in ReCALL. The year with the highest percentage of related articles was 2012 (10.3%), while no related articles were published in 2004 and 2007. Figure 1 illustrates the increasing trend of articles related to VR in the four journals from 2004 to 2013.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>0/13 (0)</td>
<td>0/14 (0)</td>
<td>0/12 (0)</td>
<td>0/13 (0)</td>
<td>0/12 (0)</td>
<td>0/12 (0)</td>
<td>0/10 (0)</td>
<td>1/12 (8.3)</td>
<td>0/11 (0)</td>
<td>0/19 (0)</td>
<td>1/128 (0.8)</td>
</tr>
<tr>
<td>2005</td>
<td>0/21 (0)</td>
<td>2/24 (0)</td>
<td>1/24 (4.1)</td>
<td>0/22 (0)</td>
<td>0/29 (0)</td>
<td>2/34 (6.0)</td>
<td>0/31 (0)</td>
<td>0/26 (0)</td>
<td>0/20 (0)</td>
<td>0/262 (1.9)</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>0/21 (0)</td>
<td>1/22 (4.5)</td>
<td>1/19 (5.2)</td>
<td>0/25 (0)</td>
<td>1/27 (3.7)</td>
<td>2/23 (8.7)</td>
<td>0/24 (0)</td>
<td>2/25 (8)</td>
<td>2/23 (8.7)</td>
<td>2/24 (8.3)</td>
<td>11/233 (4.7)</td>
</tr>
<tr>
<td>2007</td>
<td>0/34 (0)</td>
<td>0/16 (0)</td>
<td>0/12 (0)</td>
<td>0/18 (0)</td>
<td>0/20 (0)</td>
<td>1/18 (5.6)</td>
<td>1/20 (5)</td>
<td>0/13 (0)</td>
<td>6/17 (35.2)</td>
<td>4/20 (20)</td>
<td>12/188 (6.3)</td>
</tr>
<tr>
<td>2008</td>
<td>0/89 (0)</td>
<td>3/76 (3.9)</td>
<td>2/67 (3)</td>
<td>0/78 (0)</td>
<td>1/88 (1.1)</td>
<td>5/87 (5.7)</td>
<td>1/85 (1.2)</td>
<td>3/81 (3.7)</td>
<td>8/77 (10.3)</td>
<td>6/83 (7.2)</td>
<td>29/811 (3.6)</td>
</tr>
<tr>
<td>2009</td>
<td>0/36 (0)</td>
<td>3/67 (4.5)</td>
<td>1/43 (2.4)</td>
<td>0/78 (0)</td>
<td>1/88 (1.1)</td>
<td>5/87 (5.7)</td>
<td>1/85 (1.2)</td>
<td>3/81 (3.7)</td>
<td>8/77 (10.3)</td>
<td>6/83 (7.2)</td>
<td>29/811 (3.6)</td>
</tr>
<tr>
<td>2010</td>
<td>0/21 (0)</td>
<td>2/24 (0)</td>
<td>1/24 (4.1)</td>
<td>0/22 (0)</td>
<td>0/29 (0)</td>
<td>2/34 (6.0)</td>
<td>0/31 (0)</td>
<td>0/26 (0)</td>
<td>0/20 (0)</td>
<td>0/262 (1.9)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>0/21 (0)</td>
<td>1/22 (4.5)</td>
<td>1/19 (5.2)</td>
<td>0/25 (0)</td>
<td>1/27 (3.7)</td>
<td>2/23 (8.7)</td>
<td>0/24 (0)</td>
<td>2/25 (8)</td>
<td>2/23 (8.7)</td>
<td>2/24 (8.3)</td>
<td>11/233 (4.7)</td>
</tr>
<tr>
<td>2012</td>
<td>0/34 (0)</td>
<td>0/16 (0)</td>
<td>0/12 (0)</td>
<td>0/18 (0)</td>
<td>0/20 (0)</td>
<td>1/18 (5.6)</td>
<td>1/20 (5)</td>
<td>0/13 (0)</td>
<td>6/17 (35.2)</td>
<td>4/20 (20)</td>
<td>12/188 (6.3)</td>
</tr>
<tr>
<td>2013</td>
<td>0/89 (0)</td>
<td>3/76 (3.9)</td>
<td>2/67 (3)</td>
<td>0/78 (0)</td>
<td>1/88 (1.1)</td>
<td>5/87 (5.7)</td>
<td>1/85 (1.2)</td>
<td>3/81 (3.7)</td>
<td>8/77 (10.3)</td>
<td>6/83 (7.2)</td>
<td>29/811 (3.6)</td>
</tr>
</tbody>
</table>

Figure 1. Publication trends from 2004 to 2013

489
Research topic analysis

As presented in Table 2, the most common research topic was learner differences \((n = 27)\), followed by learning task \((n = 9)\). On the other hand, the most common research topic subcategory was interactive communication \((n = 12)\), followed by learner behaviors, affections, and beliefs \((n = 9)\) and TBI \((n = 6)\). The least common research topic subcategory was collaborative learning \((n = 1)\).

<table>
<thead>
<tr>
<th>Table 2. Thematic foci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>A. Learner differences</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>B. Learning task</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>C. Impact of the teacher</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>D. Environment</td>
</tr>
</tbody>
</table>

Note. In order to provide in-depth discussions, studies covered two or more categories were identified, and so the total frequency counts did not match the number of the reviewed articles.

An analysis of technologies used

The content analysis revealed that the technology used most often was open social virtualities \((65.6\%)\), while the least common were SIE \((3.4\%)\) and those in the “not specified” category \((3.4\%)\) (see Table 3). Research technologies used in open social virtualities from 2004 to 2008 employed various platforms, such as 2D text-based MOO \((n = 2)\) and AW \((n = 2)\). From 2009 to 2013, SL \((n = 12)\) became the most utilized platform. AW, Blue Mars Lite, and Wonderland were each used once only. The results also showed that researchers demonstrated more interest in the MMOG type of VR platforms, such as WoW \((n = 3)\) and Civilization \((n = 1)\). While only one platform, Quest Atlantis, was classified as an SIE, four platforms were identified as “others.” These platforms in the “others” category can be divided into 3D computer graphic/game development programs (e.g., Unity and Autodesk 3Ds MAX) and life simulation video games without a specified multiuser function (e.g., The Sims and Sim Theme Park).

<table>
<thead>
<tr>
<th>Table 3. Percentage of articles related to specific MUVE technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>Open social virtualities</td>
</tr>
<tr>
<td>MMOGs</td>
</tr>
<tr>
<td>SIEs</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td>Not specified</td>
</tr>
</tbody>
</table>

The trends of VR in research methodology

RQ4 was answered by analyzing the language learning settings, sample groups, and methodological approaches (see Table 4). It is believed that dividing the analysis of applied methodologies into two five-year-periods (i.e., 2004–2008 and 2009–2013) will provide a better reference to help researchers make plans in the future. Therefore, the changes between 2004–2008 and 2009–2013 are particularly illustrated in this section. From 2004 to 2008, the language learning settings for language exchange \((n = 2)\) and FL \((n = 2)\) were utilized most, followed by L2 \((n = 1)\), “others” \((n = 1)\), and L1 \((n = 0)\). From 2009 to 2013, the most common language learning setting was FL \((n = 16)\), followed by “others” \((n = 6)\) and L2 \((n = 1)\). The number of articles on L1 learning remained 0 in second period, and on language exchange dropped to 0. Furthermore, the percentage of articles on FL increased from 33% to 70% between the two periods, and the language learning formats in the “others” category also increased from 17% to...
26%. It is noted that the language learning formats in the “others” category were the target language for special purposes (e.g., business, architecture, and cultural awareness) or a combination of FL, L2, and L1.

<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Language learning settings/participants</th>
<th>Methodological approach</th>
<th>Data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Hansson</td>
<td>FL. 10 7th-grade pupils</td>
<td>Qualitative</td>
<td>Text chats</td>
</tr>
<tr>
<td>O’Rourke</td>
<td>Language exchange. 34 German and 24 Irish college students</td>
<td>Qualitative</td>
<td>Text chats</td>
<td></td>
</tr>
<tr>
<td>Rilling et al.</td>
<td>Others. 4 preservice language teachers</td>
<td>Qualitative</td>
<td>Learner feedback, teachers’ self-reflection</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>Peterson</td>
<td>FL. 24 college English learners in Japan</td>
<td>Qualitative</td>
<td>Text and audio transcripts, observation, questionnaires</td>
</tr>
<tr>
<td>Schwienhorst &amp; Borgia</td>
<td>Language exchange. 34 German and 26 Irish college students in 2000-2001. 18 German and 12 Irish college students in 2002-2003</td>
<td>Quantitative</td>
<td>Text chats, course statistical data</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>Ranalli</td>
<td>L2. 9 intermediate-level English learners at a university</td>
<td>Quan.+qual.</td>
<td>Testing, questionnaire</td>
</tr>
<tr>
<td>2009</td>
<td>Deutschmann et al.</td>
<td>FL. A comparison of 2 oral proficiency courses aimed at doctoral students</td>
<td>Quan.+qual.</td>
<td>Audio recordings, questionnaires, group evaluations, personal interviews</td>
</tr>
<tr>
<td>Ho et al.</td>
<td>L2. 45 12th-grade students</td>
<td>Qualitative</td>
<td>Screen recordings, questionnaires, interviews</td>
<td></td>
</tr>
<tr>
<td>O’Brien et al.</td>
<td>FL. 1st semester: 42 German high-school students  FL. 3rd semester: 33 German high-school students</td>
<td>Quantitative</td>
<td>Questionnaires</td>
<td></td>
</tr>
<tr>
<td>Peterson</td>
<td>FL. 14 college English learners in Japan</td>
<td>Qualitative</td>
<td>Questionnaire, text chats, observations</td>
<td></td>
</tr>
<tr>
<td>Zheng et al.</td>
<td>FL. 61 middle-school students in China</td>
<td>Quantitative</td>
<td>Testing, survey</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>Peterson</td>
<td>FL. 7 college English learners in Japan</td>
<td>Qualitative</td>
<td>Questionnaire, text chats</td>
</tr>
<tr>
<td>2011</td>
<td>Collentine</td>
<td>FL. 58 college Spanish learners</td>
<td>Quan.+qual.</td>
<td>Text chats, user-tracking data (learner’s movements, actions, and choices)</td>
</tr>
<tr>
<td>Jauregi et al.</td>
<td>FL. 2 Spanish learners &amp; 2 preservice teachers in universities</td>
<td>Qualitative</td>
<td>Verbal transcripts and interaction data, questionnaires, interviews</td>
<td></td>
</tr>
<tr>
<td>Wehner et al.</td>
<td>FL. 40 Spanish learners at 1 university (SL, n = 20; non-SL, n = 20)</td>
<td>Quantitative</td>
<td>Survey</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Cornillie et al.</td>
<td>FL. 83 first-year university students and learners in high school</td>
<td>Quan.+qual.</td>
<td>Questionnaires, interviews, game logs</td>
</tr>
<tr>
<td>Liang</td>
<td>FL. 11 college English learners</td>
<td>Quan.+qual.</td>
<td>Students’ verbal and chat transcripts</td>
<td></td>
</tr>
<tr>
<td>Liou</td>
<td>FL. 25 potential English teachers</td>
<td>Qualitative</td>
<td>Questionnaires, interviews</td>
<td></td>
</tr>
<tr>
<td>Peterson(a)</td>
<td>FL. 8 college English learners in Japan</td>
<td>Qualitative</td>
<td>Text chats, observation, interviews, questionnaire</td>
<td></td>
</tr>
<tr>
<td>Peterson(b)</td>
<td>FL. 4 college English learners in Japan</td>
<td>Qualitative</td>
<td>Text chats, observation, questionnaires, interviews</td>
<td></td>
</tr>
<tr>
<td>Rama et al.</td>
<td>FL. 6 college English learners</td>
<td>Qualitative</td>
<td>Observation, interviews, chat</td>
<td></td>
</tr>
</tbody>
</table>
Table 4 indicates that from 2004 to 2008 the most common research sample groups were in higher education \( (n = 4) \), followed by teachers \( (n = 1) \) and elementary school \( (n = 1) \). There were no research groups in the junior/senior high school, adults, special needs, and “others” categories during the initial period. From 2009 to 2013, research samples in higher education were still used for most of the VR research papers \( (n = 13) \), followed by “others” \( (n = 4) \), junior/senior high school \( (n = 3) \), teachers \( (n = 2) \), and adults \( (n = 1) \), with none for elementary school \( (n = 0) \) and special needs \( (n = 0) \). Furthermore, declining trends were evident between the two periods in the elementary school (from 17% to 0%), higher education (from 66% to 57%), and teachers (from 17% to 9%) research subcategories, while researchers demonstrated increased interest in “others” (from 0% to 17%), junior/senior high school (from 0% to 13%), and adults (from 0% to 4%).

The results indicated that from 2004 to 2008, the quantitative methodological approach was found in most of the VR publications \( (n = 3) \), followed by the quan.+qual. \( (n = 2) \) and quantitative \( (n = 1) \) approaches. From 2009 to 2013, the top ranked research design was still qualitative \( (n = 13) \), followed by quan.+qual. \( (n = 6) \), and quantitative \( (n = 4) \). The research interest in quan.+qual. decreased from 33% to 26% between the two periods, while that in the qualitative methodological approach increased from 50% to 57%, while the quantitative research approach remained at 17%.

**Discussion**

This study aimed to identify the trends in VR research articles in terms of their research topics, technologies used, language learning settings, sample groups, and methodological approaches. The results are discussed below.

**VR trends in selected journals and research topics**

The findings show that the number of VR studies of language learning gradually increased from 2004 to 2013. This is consistent with Hsu et al. (2012) reporting that the number of articles on technology-based learning concerning “digital games and intelligent toy enhanced learning” increased significantly between 2000–2004 and 2005–2009 (from 0.81% to 3.82%).

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Sample Group</th>
<th>Methodology</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>Canto et al.</td>
<td>FL, 36 Spanish learners at the University of Utrecht and 14 preservice teachers at the University of Valencia</td>
<td>Quantitative</td>
<td>Observation, verbal testing, questionnaires</td>
</tr>
<tr>
<td></td>
<td>Thorne et al.</td>
<td>Others. 32 Dutch &amp; 32 American WoW gamers</td>
<td>Qualitative</td>
<td>Verbal transcripts and interaction data, observation</td>
</tr>
<tr>
<td></td>
<td>Zheng et al.</td>
<td>Others. Three college-age English learners</td>
<td>Qualitative</td>
<td>In-/out-of-game texts, interviews, questionnaire</td>
</tr>
<tr>
<td></td>
<td>Dooly &amp; Sadler</td>
<td>Others. Student-teachers studying to become language teachers in Spain and the USA</td>
<td>Qualitative</td>
<td>Classroom and screen recordings, verbal and text transcripts, forum posts, email exchanges, self-evaluation, field notes, etc.</td>
</tr>
<tr>
<td></td>
<td>Ryu</td>
<td>Others. 6 adults whose L1 is not English</td>
<td>Qualitative</td>
<td>Interaction data, observation, interviews</td>
</tr>
<tr>
<td></td>
<td>Shih</td>
<td>FL, 4 college English learners in Taiwan</td>
<td>Quan.+qual.</td>
<td>Cultural knowledge tests, interviews, observation, blog entries</td>
</tr>
<tr>
<td></td>
<td>Wigham &amp; Chanier(a)</td>
<td>Others. 17 architecture college students</td>
<td>Qualitative</td>
<td>Screen recordings, observation, verbal chats, questionnaires</td>
</tr>
<tr>
<td></td>
<td>Wigham &amp; Chanier(b)</td>
<td>Others. 17 architecture college students</td>
<td>Quantitative</td>
<td>Screen recordings, text and audio transcripts</td>
</tr>
</tbody>
</table>
The content analysis revealed that the most common research topic from 2004 to 2013 was learner differences—interactive communication. Previous VR research has investigated the role of metalinguistic awareness (O’Rourke, 2005), the kind of interaction management strategies employed by learners (Peterson, 2006), and the factors influencing the use of these strategies in VLEs (Peterson, 2009). Based on this valuable groundwork, some follow-up studies have suggested that 3D VLEs affect learner behaviors in terms of linguistic complexity, accuracy, and correct feedback (e.g., Thorne et al., 2012; Wigham & Chanier, 2013b). As more 3D technologies were developed, the use of nonverbal communication (e.g., the modalities of proxemics, kinesics, and avatar appearance) has been highlighted to overcome verbal miscommunication in virtual worlds (Wigham & Chanier, 2013a). It is expected that future research will focus on examining the effectiveness of learner interactive communication in different VR contexts.

Additionally, learner behaviors, affections, and beliefs have also received considerable attention between the two 5-year periods. The related studies have consistently found that VLEs increase learner autonomy and self-efficacy, reduce learning anxiety, and foster creativity. VR presents a realistic virtual space and visible “classmates” who assist students in gaining a sense of participation and building emotional bonds (positive or negative) with their collaborative partners (e.g., Collentine, 2011; Deutschmann et al., 2009; Peterson, 2012b). In contrast, negative perspectives have also been reported, mostly associated with the use of specific VR tools that had user-side and server-side issues. Overall, a growing trend of individual differences and preferences regarding social identities and metacognition in virtual worlds should be expected as more learner control functions become available with the use of these latest advanced technologies (Hsu et al., 2012).

The topics in the learning task—TBI category received particular attention from VR researchers between the two periods. TBI is characterized by activities that are generally theme venues and emphasize engaging learners in meaningful, goal-directed communication to collaboratively solve problems and complete assigned tasks. Several studies found that conducting TBI in VLEs, which involves authenticity and collaborative elements, has a direct impact on learner participation, engagement, and the amount of negotiation (e.g., Peterson, 2012a). Since TBI has a wide range of theoretical groundings (e.g., interactionist theory, sociocultural theories, and experiential learning theory), more research into integrating TBI in VR is expected.

The use of VR tools in CALL

Open social virtualities was found to be the most popular research VR tool, with a consistently high level of interest across the selected studies. Earlier VR research focused more on MOO and AW, whereas studies performed between 2009 and 2013 switched to the use of SL. This finding suggests that text-based 2D VLEs are no longer able to satisfy the needs of practitioners. Rather, 3D VR tools with the features of multiple communication channels and high visual appeal are better suited to creating tasks with greater degrees of interaction and collaboration, and thus can provide multiple modalities for input, output, and feedback. As a consequence of being widely investigated in the CALL field, we suggest that SL provides more useful features for language learning than do other open social virtualities (Warburton, 2009). Additionally, the ReCALL special issue in 2012 prompted a sudden increase in the use of MMOG platforms, although this was followed by only one study in 2013; this implies that we are still in the exploration stage of MMOG technology in the CALL field. Future trends should be tracked to see how MMOGs can be used for language education in the future.

Notwithstanding the positive research results mentioned above, several negative attitudes and both user-side and server-side problems were mentioned throughout most of the studies. In terms of user-related issues, novice SL users seem to consistently report negative experiences associated with significant investments of time and energy, complicated in-world interfaces, and difficulty in writing scripts and modeling behaviors. Server-side issues included downtime, frequent updates, lag, and large monetary investments (if wishing to be more than a mere spectator) and hardware requirements (e.g., good graphic cards and a high-speed Internet connection) (e.g., Ho et al., 2009; Peterson, 2011). Fortunately, some VR software companies have been collaborating with other companies that operate in different fields of arts and sciences (e.g., topography and physics) in the development of 3D VLEs in order to make them more user-friendly. Also, the suppliers of VR platforms such as SL have offered cost discounts for educational and nonprofit institutions. User-related issues could also be solved through long-term training by joining education-focused online communities, such as the EUROCALL/CALICO Joint Virtual Worlds Special Interest Group (http://www.eurocall-languages.org/sigs/vw) and the Virtual Worlds Education Roundtable.
However, these associations are more focused on teachers than on learners. As Hauck and Youngs (2008) claimed, there is a high degree of risk in assuming that students will automatically be familiar with the affordances provided by an online environment. Several studies have claimed that it is essential to provide students with well-organized prestudy technical training to ensure that in-world learning will be effective.

The trends of VR in language learning settings, sample groups, and research designs

The FL setting was found to be the most common in the VR publications analyzed. This is not surprising, given that FL learners often do not have ready access to a suitable environment in which to practice and use the target language. VLEs can overcome this difficulty by providing an immersive and authentic environment to socially interact with native speakers. The data also showed that the “others” category was the second most common research language learning setting, appearing in four out of six articles published in 2013. This implies that language learning environments are becoming more diverse. The learning settings available in cyberspace cannot be divided into traditional ones, such as FL, L2, or L1; rather, a virtual environment can be a combination of all other environments, in which users are not constrained to their physical locations—they can mingle together without physical frontiers.

It is noteworthy that two of the earlier studies investigated whether the use of the MOO medium and tandem exchange—based on mutual language exchange between partners who are native speakers of the target languages—affect on the negotiation of meaning (O’Rourke, 2005), and whether the implementation of MOO improves the balance in bilingualism (Schwienhorst & Borgia, 2006). Surprisingly, this language learning format has not received much attention in recent years, even though the combination of VLEs with tandem courses would make it easy to create language partnerships across countries.

The current study also found that higher education was the sample group utilized most in VR publications, with sample groups in elementary and junior/senior high schools being explored less frequently over the last 10 years. Several factors could explain this research trend. First, this result reflects Rankin, Gold, and Gooch (2006) finding that lower-level language learners experience difficulties dealing with multiple competencies required by the 3D VLEs and can experience cognitive overload. Second, some VLEs have age restrictions, such as 16+ years for SL and 13+ years for WoW. Most VR platforms are commercial products, and the instructors have little control over who interacts with students outside class times and what locations they visit in a virtual world. Safety therefore remains a concern, especially in K-12 education settings. To avoid these problems, SL has divided its regions into general, moderate, and adult regions to avoid adult content being available to users younger than 18 years. Additionally, some software programs have been developed for use by people of all ages by carefully incorporating pedagogy into the immersive spaces (e.g., AW Educational Universe and Croquelandia).

Finally, the findings reveal that the qualitative methodological approach was dominant in the VR studies, and that nearly one-third of the articles used quan.+qual. techniques. This is a reflection of the complexities involved and the multiple sources of the problems addressed in these studies. Regarding the instrumentation used for data collection, the relevant studies performed over the past 5 years have overcome earlier research deficiencies, such as solely utilizing text chats or transcribed audio tapes, which had resulted in a lack of nonlinguistic and paralinguistic information (O’Rourke, 2005). As presented in Table 4, more recent studies have employed screen recordings or embedded user-tracking systems with triangulated data collection to strengthen their credibility (e.g., Collentine, 2011; Wigham & Chanier, 2013a). Spector et al. (2014) reported that it has become common to employ qualitative or quan.+qual. methods with triangulation techniques to investigate and explain complicated issues encountered in the field of education technology research. Regarding the procedure of data collection, recent attention has emphasized at-home tasks (Liou, 2012) and the beyond-game culture (Ryu, 2013) for language learning in virtual worlds, rather than multichannel in-class data collection. This is consistent with Peterson (2012b) predicting that there will be an increase in the number of education studies focusing on informal outside-class language learning rather than on formal in-class language learning. This situation has resulted in the recruitment of volunteers as research participants directly in VLEs becoming a new trend of data collection (see Ryu, 2013).
Conclusion and future prospects

This study investigated trends in VR research from 2004 to 2013 in the four top CALL journals. The research topics that have been widely investigated during this period (interactive communication; behaviors, affections, and beliefs; and TBI) are consistent with VR being advocated as a promising arena for language learners. The findings reveal that FL educators in higher education are now widely expected to deliver their language instruction with the aid of VR technologies. Regarding the methodologies, qualitative and quan.+qual. methods with triangulated data collection will continue to provide rigorous and in-depth analysis of language learning in VLEs. We believe more advanced and new methods of data collection, such as brain activity and eye tracker data, as well as informal language learning will evolve to more accurately capture in-world language activities.

However, the content analysis performed in this study revealed several areas in which considerable efforts are urgently needed. First, only a small proportion of the included articles examined teachers’ perspectives and the awareness of task design. As Peterson (2011) pointed out, a most effective role of the teacher in VLEs has yet to be clarified. As a result, little is known from empirical research about how instructors’ roles change in a VR classroom, teachers’ decision-making on how to integrate pedagogical activities into VLEs by utilizing the strengths of VR, and how to motivate teachers to adopt and continue using VLEs when teaching (e.g., the method of teacher preparation). Second, regarding the less developed language learning settings and sample groups, studies of language for specific purposes and the tandem principle of bilingualism are worth exploring further. As mentioned above, the TBI approach in theme-based venues can facilitate bilingual language exchanges and language learning for specific purposes (e.g., tourism and medical applications). The positive results obtained when using VR are consistent with 3D VLEs having the advantage of multichannel communication, which can effectively reduce learning barriers (e.g., anxiety and inhibition) (Wehner et al., 2011). Thus, it was felt that virtual worlds have particular potential in assisting language learners with special needs. For example, students with autism may benefit from learning languages in virtual worlds since these students might respond better in a low-stress environment where they have more time to master tasks. We believe that the continuing explorations in the VR field will lead to the development of valuable new perspectives about and methods of language development.

Finally, this study did not consider relevant conference proceedings and other language and technology-related publications. Future research should expand the data sources, as well as consider other journals and other analysis techniques (e.g., citation analysis) so as to include more of the available literature.

Acknowledgements

Funding for this research work is provided by the Ministry of Science and Technology, Taiwan, under Grant Nos. NSC 101-2511-S-003-031-MY3, MOST 104-2911-I-003-301, and MOST 103-2628-S-003-002-MY3. This research is also partially supported by the Aim for the Top University Project and Center of Learning Technology for Chinese of NTNU, sponsored by the Ministry of Education, Taiwan, and the International Research-Intensive Center of Excellence Program of NTNU.

References


Acceptability and Satisfaction of an ICT-based Training for University Teachers

Rocío Herrero¹, Juana Bretón-López¹,³, Luis Farfallini¹*, Soledad Quero¹,³, Ignacio Miralles¹,³, Rosa Baños²,³ and Cristina Botella¹,³

¹Jaume I University, Castellón de la Plana, Spain // ²University of Valencia, Valencia, Spain // ³CIBER Pathophysiology Obesity and Nutrition, Institute of Health Carlos III, University Hospital of Santiago de Compostela Choupana, Spain // ro.herrero.09@gmail.com // breton@uji.es // lfarfallini@gmail.com // squero@uji.es // mirallei@uji.es // banos@uv.es // botella@uji.es

*Corresponding author

(Submitted March 9, 2014; Revised January 5, 2015; Accepted March 20, 2015)

ABSTRACT

E-learning can be defined as learning facilitated and supported through the use of ICTs. ICTs can increase students’ motivation, accelerate the knowledge process and facilitate the information access. The aim of this paper is to analyze the acceptability of three tools presented in a workshop carried out during six weeks, with university teachers (N = 22), taking into account their opinion about the usefulness of the workshop. Two researchers/teachers specializing in the field of educational technologies taught the three tools presented. Before the workshop, we administered a technology profile questionnaire; after it, we asked the teachers to fill in a questionnaire about the workshop and the tools used. We then carried out a non-parametric Friedman test to compare the differences found in the evaluation of the tools. The results show that the acceptability of the technologies and the methodology used during the workshop was high: teachers’ assessment and opinion has been favorable for both pedagogy and teaching methodology, which facilitated the understanding of the tools and the innovative nature of the contents of the workshop. The participants also evaluated the technologies as appropriate and easy to use. We also mention the limitations of the study and future challenges.

Keywords

Interactive learning environments, Lifelong learning, Teaching and learning strategies, Virtual reality, Information and communication technologies

Introduction

For more than twenty years, the U.S. Department of Education has been proposing an educational change in order to create a more effective system. In this sense, the field of affective neuroscience has drawn attention to the critical importance of motivation in how the brain learns (Wolfe & Brandt, 1998). Motivation is a key concept in the academic performance (Porter, Bigley, & Steers, 2003). People learn to pay attention on what interests them, which can vary from learner to learner. Each student wishes to know and acquire a specific range of knowledge, skills, and competences. Therefore, an effective learning experience has to be individualized to satisfy the interest of each learner. In their recommendations, the U.S. Department of Education proposed an education program mediated by a set of learning tools designed to personalize the learning. In this sense, the use of educational technologies is suggested as an important innovation to enhance the current system and achieve this goal (U.S. Department of Education, 2010a, 2010b; Uzumboylu, 2006).

Everywhere we look, information and communications technologies (ICT) are increasingly surrounding us. They are inside every area of our life, generating several changes in the way that we behave. In this sense, education is not an exception, and the change in education has come from two directions. On one hand, the changes generated in education come from outside, generated by demands from the society, as with, for example, those companies seeking for skilled workers. A decade ago, the instruction on technologies and the expertise in the use of these technologies were considered an advantage in which a candidate could choose to increase the importance of his or her education. Nowadays, this knowledge is mandatory, a prerequisite, where those in low-qualification jobs still need this instruction to carry out even basic tasks. However, on the other hand, the demand also come from inside. The relevance of technologies being introduced in the educational field is not only about updating the knowledge needed by the students for their future or using the potential that technology adds. It is also about speaking in the same language that students speak. We live in a highly mobile, globally connected society in which the younger people have grown up as “digital natives” (Hansen, 2003; Prensky, 2001, 2009). As digital natives, the way in which they
interact with information and communicate between each other is very different from the ways that the traditional schools propose (Bennett, Maton & Kervin, 2008).

In this sense we need to prepare students for this new world. We must change what and how we teach in order to match what students need to know and where and when they learn. We need to bring technology into learning in meaningful ways so that we engage and motivate learners to achieve the required knowledge and expertise.

In the present day, the education system depends on the relationship between teachers and their students, and it mainly supports learning from textbooks and, in some cases, through the use of technologies, although in a quite limited way. This is mainly based on the use of videos, PowerPoint presentations, or through a limited use of the Internet. However, technology also provides access to a much wider and more flexible set of learning resources. Nowadays, we have different technological devices that enable and enhance communication and knowledge transfers. Their potential utility in the field of education is considerable. Moreover, the challenge to the education system is to incorporate technology to create relevant learning experiences that mirror students’ daily lives and the reality of their futures (Bransford et al., 2006). The use of technologies is not generally well exploited or at least not fully exploited.

E-learning can be defined as learning facilitated and supported through the use of ICTs. The term “e-learning” therefore essentially covers the use of computers and technology as a vehicle for knowledge exchange within teaching and learning. It is a recent concept (Sun, Tsai, Finger, Chen, & Yeh, 2008) that has emerged from the new paradigm of modern education. ICTs offer very useful tools for teaching and learning processes. Indeed, ICTs include a variety of applications such as the Internet, virtual environments, and serious games that can be used to improve some important skills, knowledge, and contents. These tools have several proven advantages, including an increase of motivation in the students (Papastergiou, 2009), a broader connection with reality, accelerating and improving the process of knowledge (Ponce, Mayer, & Lopez, 2013) and facilitating access to the information (Thompson, 2013). They proved to be highly effective in fostering skills, among which collaborative work stood out as prominent (Di Blas & Paolini, 2014). Also, they had some disadvantages such as problems related to distraction, including poor academic performance and comprehension (Fried, 2008; Sana, Weston, & Cepeda, 2013). In this line, Kong et al. (2014) identified some research issues critical for e-learning: the developing of some skills (communication and collaboration skills) and the maximization of learning opportunities during the learning process, among others. It is important to understand how these technologies work and under what circumstances they can be used to improve the learning process and also under what circumstances they are not recommended.

The concept of e-learning is growing in importance. Proof of this lies is the interest shown by international organizations such as the European Commission, which has increased the economical resources available on the development and generation of programs, including ICTs in the educational field. One example is the Teaching to Teach with Technology (T3) KA3 ICT Multilateral Projects (PROJECT NUMBER 505169-LLP-1-2009-1-IT-KA3-KA3MP).

The main goal of the T3 Project was to develop and validate an innovative teaching program to promote the use of e-learning tools in different contexts (Bretón-López, Botella, Vizcaíno, Quero, Baños, & Molés, 2010; Bretón-López, Quero, Botella, Baños, Farfallini, & Herrero, 2011; Bretón-López, Quero, Botella, Baños, Vizcaíno, Farfallini & Herrero, 2011). T3 comprised three trials in different countries (United Kingdom, Spain, and Italy) using a variety of technologies (web-based, Internet-based training, and virtual learning environments) in diverse educational contexts (secondary schools, universities, and commercial companies). The final purpose of T3 is to involve teachers directly in developing novel IT-based teaching practices. In this sense, the dissemination tools proposed by the project are designed to inspire, encouraging teachers to develop their own innovative applications to improve their own practice. The teachers’ beliefs play a critical role when we include the technology in their classes and also influence in the acceptability of these tools, but few studies have examined those beliefs (Howard, 2011).

This paper aims to show the results of acceptability and satisfaction of Spanish teachers with three software programs presented in a workshop of ICT-based learning technologies for education in a university context, within the framework of the European project T3. The sample was composed of university teachers from different disciplines who were interested in the use of innovative teaching technologies. We present relevant data about teacher satisfaction with the technologies used and their opinion about the relevance of incorporating ICT-based learning tools in a university context.
Material and methods

Participants

The recruitment of the participants was conducted through the university email list. Information about the workshop was sent to teachers who usually participate in university training courses. The teachers interested in the workshop requested admission. Initially, 22 people registered. Two participants failed to start the course because of scheduling difficulties, and four did not meet the minimum attendance required for consideration in the final sample. Finally, the group was composed of 16 participants: 10 women and 6 men, with a mean age of 32 and a standard deviation of 4.4. Participants had various different levels of teaching experience, ranging from one to eight years’ teaching at the university.

The participants had different academic backgrounds: five were psychologists, six were engineers, two were chemists, one was a translator, one specialized in information science and publicity, and one had a computing degree.

All of the participants were teachers at Universitat Jaume I (UJI), but with different types of contracts: Six were pre-doctoral fellows, four were assistant professors, three were associate professors, two were contracted research staff, and one was a member of the research teaching personnel.

Selected technology and reasons

Literature shows us that educational research has generated many methodologies, tools, and practices exploiting the potential of technology to improve education (Jin & Bridges, 2014). Games are not an exception and there are several examples where, appealing to entertainment, people can learn new things (Butler, Someya, & Fukuhara, 2014) and even change their habits of life (Baños et al., 2013). However, it is important to underline that the education sector and also teachers need a better understanding of the potential and the diversity of such tools (Sica, Nigrelli, Rega, & Miglino, 2011). The different learning contexts in which the educational relationship takes place (Olson & Bruner, 1974) use different means of transferring knowledge and need to be calibrated. For example, the use of serious games is particularly appropriate for young people because while they are playing they are also learning (Laudon & Laudon, 2007), but not for adults, given that they are accustomed to a different type of training (Sefton-Green, 2006). Therefore, the kind of tool selected to be used in each learning context is an important point to consider.

In order to make a selection, a classification of the learning technologies was carried out. The resulting categories were based on the new trends in educational psychology (Jessel, 2011) and took two variables into account: (1) the type of technology (based on instructions or constructivism), and (2) the type of teaching (experimenting, experiencing soft skills, or exploring).

Category 1 refers to technologies based on instructions, which develop environments to create educational materials that can be used even by non-computer experts. Constructivist technologies are self-contained applications that propose activities in a particular domain and use precise specifications. Category 2 refers to the main teaching practice. Sica et al. (2011) point out three types of teaching and learning strategies involved in games experimenting, experiencing soft skills, and exploring:

- The demonstration-experiment is one of the most traditional strategies used by teachers. For example, performing laboratory experiments in which the student is involved in his or her own experience.
- Learning by experience is an explicit learning focused on the working environment. It is both individual and collective and is focused not on knowledge, but on skills, attitudes, and expertise. The learner has an active role and consciously learns through collaboration with others and under the guidance of experts in safe environments.
- Exploring is an innate human propensity to experience the environments in which they are to act. Many educational practices used this tendency to transfer their skills and knowledge. The adventure game is an example of this type of educational practice, where the transpositions to the technological environment take place.

According to this classification, and taking into account the university context in relation to the relevance of the global origins of their students (especially since the Bologna process began), we considered that experiencing and
exploring soft skills would be a relevant topic for the workshop. Also, considering the model explained by Jessel (2011) in terms of classification of learning technologies, a constructive approach is the best way to teach this material. Three technologies were selected as the most adequate ones to achieve our goals: e-Adventure, Eutopia, and PalMa systems. Other criteria followed during the selection were that technologies had to belong to an open educational resource, reside in the public domain, and have an intellectual property licence in order to allow participants the use of the technologies in their future educational practices.

**e-Adventure**

The e-Adventure platform is the result of a research project aimed at facilitating the integration of educational games and game-like simulations within educational processes this is being developed by the e-learning research group at Universidad Complutense of Madrid. e-Adventure is a platform for the development of classic adventure computer games with an educational slant.

The e-Adventure engine includes a built-in assessment mechanism that can be used to automatically grade the student or generate human-readable reports to be processed by instructors for assessment purposes. Once the game is completed, the assessment report is generated. The
instructors can access the results via the web, and the information can be shown to the students. The games are encapsulated as a learning object with standardized metadata that allows their storage and discovery in standards-based repositories of learning content.

More details about the tool can be found at http://e-adventure.e-ucm.es/.

**Europia**

Europia is a platform designed to support distance learning. It is a useful platform to create and organize educational multiplayer online role-playing games. The term role-play describes a range of activities characterized by involving participants in as-if or simulated actions and circumstances that project into an imaginative-creative process established through the interpretation of a real or fictional role in a specific given situation. Role-play has been extensively recognized as a powerful technique for enhancing the traditional training practice, boosting participants’ learning experience, facilitating knowledge, and promoting skills and competencies in groups, as well as for personal development. This approach allows a small group of people to give a theatrical performance for educational purposes. Its main intrinsic value is to be a flexible method that allows participants to experience realistic learning scenarios in a way that best suits specific needs, situations, and learning styles.

Europia is an online role-play platform conceived and designed around the presence of a group of players interacting with each other through the presence of a digital alter ego (an avatar) and under the supervision and guidance of a role-play director, because we have seen these as key aspects for meaningful learning experiences.

![Figure 2. Eutopia editor](image)

Each actor (or learner) is represented by an avatar that interacts with other avatars controlled by real people in a virtual 3D scene (see Figure 2). The director (who, according to the context of role-play setting applications, can be a...
teacher, trainer, educator, or consultant) can play different roles. They can write a storyboard as a playwright, assign roles to players (goals, characters, and the personalities of individual avatars) as a casting director, guide the action in the performances as a movie director, and finally, they can give personalized feedback to the group by recording and analyzing a significant part of the scene of the enacted performance (feedback and debriefing phase).

Players communicate via short text messages and non-verbal communication features, like gestures, reproduction of volume, and tone of voice. Players can control avatars’ gestures and body movements. They can also whisper messages to each other. These messages are audible only to the other partner in the conversation and to the trainer. Finally, they can communicate with the trainer to ask for advice or clarification or to raise any other questions about the online simulation. Once the game is in progress, trainers can observe what is going on from either viewpoint (first-person and third-person point of view), get involved at any moment, send messages to players, or activate special events or happenings. When the game session is over, the trainer can lead a debriefing session in a group discussion, analyzing the communication and behaviour strategies adopted by the players.

In educational fields, Eutopia allows interaction in real time between students; it also allows the teacher to see the students’ interactions, give them feedback, and interact with one or all of them. More details about the tool are provided at http://www.nac.unina.it/eutopia/download.htm.

**PalMa**

Palestra Manageriale (PalMa) is a serious game conceived as a software tool with a specific learning outcome. It is a simulator of dialogues in which the human user is confronted with a bot (a software agent conceived to answer in a predefined way during the interaction) and, through a series of communicative exchanges, seeks to achieve a certain goal. The versatility of PalMa allows to train users in a wide range of soft skills such as leadership, negotiation, and effective communication skills.

The reference unit of PalMa is a scenario, a situation where the player is asked to achieve a predefined goal. The player acts through an avatar. The interlocutor of the player is a bot, in reaction to the action made by the player (see...
Figure 3). From time to time the user selects a phrase (from a number of possible options) that he or she considers effective with respect to this objective. The bot responds in turn with its own feedback. If the user adopts an effective communication strategy reaches the goal. Otherwise the dialogue fails. PalMa is particularly flexible in terms of teaching because it can simulate any kind communicative exchange that occurs between two people.

Whoever designs the training customizes the game, defining the environment in which the characters act, assigning the characters of both the user’s avatar and the bots, assigning the level of difficulty, and establishing the response options available to the player and the bots in different exchanges provided by the gym.

PalMa provides several useful feedbacks to generate learning. The first kind of feedback is related to the information entered directly into the game. PalMa adopts the graphic language of comics, inserting every sentence inside a callout. The shape, colour, and size of the text in the callout allow the user to understand if the discussion is proceeding correctly or if the sentences selected are not. In addition, the user also has available data, graphics, and comments on the players’ performance at the end of the game.

In educational fields, PalMa allows teachers to design exercises with different levels of difficulty to train and test their students in specific previously selected skills. More details about the tool are provided at http://www.entropykn.net/palma-seriousgame/.

**Measures**

The following assessment protocol was applied to the participants:

- **Technology profile questionnaire (TPQ):** This is a questionnaire designed by our research team in order to collect information about previous experience with computers and new technologies. Its aim is to build a technological profile of the different technologies used by the sample before the beginning of the workshop. This questionnaire is composed of 23 items, with a scale of response ranging from 1 (never) to 5 (very often). Each item corresponds to a specific technological tool, which the trainees could have used in their educational contexts (“Please indicate how often you use the following tools, e.g., conventional desktop or laptop; generic software tools, PowerPoint, data handling, word processing; games for educational purposes, etc.”). This tool was applied on the first day of the trials.

- **Tool Evaluation Questionnaire (TEQ):** The aim of this questionnaire is to evaluate each form of technology used in the workshop in terms of design, usefulness perceived by the user, and usability. This questionnaire is composed of 13 items (e.g., “I found the technology easy to use”; “Things I needed were visible or easy to find”; “I would recommend the technology that I used to other colleagues”), with a scale of responses ranging from 1 (strongly disagree) to 5 (strongly agree). This tool is applied on the last day of the trials, and one questionnaire is applied per each tool.

- **Workshop Evaluation Questionnaire (WEQ):** The aim of this questionnaire is to find out how helpful the training course was. The questionnaire has 11 items (“The course helped me find ways of using new technology in different learning contexts”; “The course has helped me explain the advantages of the technologies”; “I will continue to use and experiment with new technologies”; etc.), with response ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). This tool is applied on the last day of the trials.

**Workshop**

The main goal of the workshop was to introduce the use of technologies to university teachers as tools to enhance their own practice. The workshop comprised six sessions of eight hours, with the following structure (see Figure 4).

Next, the goals of each session will be explained.

The main objectives of the first session were to give a general presentation about the T3 Project and find out the different interests of the participants in relation to the application of ICTs in their own teaching practice, their knowledge about the topic, and their general experience with the new technologies. Moreover, a pre-evaluation protocol was applied; specifically a technology profile questionnaire. The aim of this first questionnaire was to find out which of the 23 technologies had been used by the trainees in their teaching contexts. The participants’ answers
could range from 1 (“never”) to 5 (“very often”). In addition, the questionnaire included items on socio-demographic data and academic background. After that, a presentation of all the tools was made.

For the second session, a theoretical introduction to soft skills was first given to contextualize the different tools that the participants would use in the workshop. After that, PalMa tools were explained and their possibilities and limitations outlined. Once participants were familiar with the program, a practical session was set up, which consisted of designing of a small example showing how it would be used. Participants applied what they learned to a schema of a hypothetical application in their own subject. After that, the participants’ experience during the practical session was discussed and conclusions were drawn regarding future uses of the program with students.

During the third session, an explanation of the Eutopia editor was given with attention being paid to its characteristics and its similarities and differences with PalMa. This was followed by a practical session that focused on the use of Eutopia. Participants were placed in different groups and, using an existing example, interacted with Eutopia. After the practical session finished, the experience of taking part in the exercise and possible applications of Eutopia were debated.

In the fourth session, we followed the same structure. First, we gave a theoretical explanation of e-Adventure tools. Once the participants became familiar with the program, a practical session was carried out using the tool. During the practical session, each participant had to follow the guidelines for constructing a game using e-Adventure. After the practical task was completed, a debate about the experience and the possible applications of the program took place.

The fifth and sixth sessions were entirely practical. The main purpose of these sessions was to give participants opportunities to design real scenarios for their own teaching. With that objective in mind, the teachers first shared their different ideas and then selected the right tools to make these possible. After the participants decided what they wanted to do, they started to build their own scenario. At the end of this sixth session, each participant shared their own project, received feedback from their partners, and completed the post-evaluation.

**Figure 4. Workshop schedule**

**Pre-evaluation**
- Technology Profile Questionnaire

**Workshop Training**
- 6 session - 8 hours each

**Post-Evaluation**
- Tools Evaluation Questionnaire
- Workshop Evaluation Questionnaire

**Procedure**

Different technologies were selected to teach the professors. An analysis of the proper methodology to teach the technology to university teachers was carried out in the context of T3 consortium meetings. It was concluded that the implementation of a workshop would be suitable in order for the university teachers to acquire the knowledge of the chosen systems, their main functions, and the specific utility for university teaching.

The workshop “the Use of New Technologies of Information and Communication for Improving Teaching” was offered to professors. The workshop was assessed by the Centre for Education and New Technologies (CENT) of the UJI. This centre is the first in Spain to improve teaching and learning through the new information and communication technologies. The CENT promotes the educational use of ICT, evaluating the teaching methodologies and technological solutions and helping professors to develop best practices using ICT. Once CENT approved, the workshop was introduced in the training courses of formation for university teachers organized by UJI. Then, the information was sent through email to teachers. Those interested in the courses responded to the email and were contacted. The workshop took place in the Educational Support Unit at UJI. Two teachers and researchers who
were specializing in the field of education and new technologies, taught the workshop. They were both psychology graduates with master’s degrees and PhD students in their teaching phase. They were both trained in the specific contents of the workshop and had more than three years’ experience in the use of new technologies in psychology.

The workshop was composed of six sessions (see Figure 4), and each one lasted eight hours. During the first and the last session, the participant must complete the pre- and post-evaluation, respectively. The pre-evaluation consisted of the TPQ to explore the previous knowledge about technologies and its frequency of use for each participant. At the end of the workshop two questionnaires were applied: the TEQ, to evaluate each form of technology used in the workshop in terms of design and usability, and the WEQ, to discover participants’ opinions about how helpful the course was in understanding new learning concepts, how innovative and useful the technologies are, and the future application of the tools.

**Results**

In order to analyze the participants’ previous experience with these technologies, we divided the information obtained by the TPQ into four different categories: traditional tools (conventional desktop, generic software, digital camera, digital audio), advanced tools (the use of robots, immersive technologies, simulations, computer modeling, hand-held technologies, virtual environments, production tools), Internet and communication tools (Internet, email, web 2.0, sharing information, communication tools, individual authoring tools, team work), and educational tools (games for educational purposes, e-portfolios, e-assessment, managed learning environments, construction of knowledge tools).

The first and highest average in the use of technologies was the Internet and communication tools category ($M = 3.61, SD = .91$); the second one, the traditional tools category ($M = 3.19, SD = 0.86$); and finally, with the same the advanced tools category ($M = 2.16, SD = .64$) and the educational tools category ($M = 2.16, SD = .80$). For a graphical view of this information, please see Figure 5. In relation to this result, it is good to know that the high level of technology use shown by the professors in their profiles may be explained by the fact that UJI encourages the use of technology very much.

![Figure 5. Technological tools used by teachers](image)

The results on the post-evaluation protocol are presented in two parts: the general evaluation of each technology as tools for their educational setting given by the participants and the participants’ opinion about the training course methodology in these technologies.

Regarding the evaluation of each technology, in general, participants considered the three tools as relevant to implement in their educational setting. According to their opinions, e-Adventure seems to be the most adequate tool
(M = 4.04, SD = .51). The second best assessed tool was Eutopia (M = 3.65, SD = .31), and finally PalMa (M = 3.32, SD = .56). For a graphical view of this information, please see Figure 6.

Also, a non-parametric Friedman test was carried out to compare those differences. Significant difference were found between the tools (X²(2) = 20.87, p < .05), which were lower in PalMa (M = 3.32) compared with Eutopia’s score (M = 3.65), and both in relation to e-Adventure’s score (M = 4.04). This result shows that e-Adventure is considered as the most appropriate tool.

Finally, regarding the opinions about the training course and the methodology used, the items were summarized in four relevant concepts: the benefits of the course for understanding, the innovation offered by the course, its usefulness, and the future applications of the acquired knowledge. The participants considered it helpful to understand new concepts of learning supported by new technologies (understanding: M = 3.81, SD = .96). They also considered they had learned innovative technologies (innovation: M = 4, SD = .98), and evaluated the technologies as useful (usefulness: M = 3.54, SD = .83). In relation to the possibilities of implementing these technologies in their future, they considered they would keep using them and would encourage others to use them as a learning method (future application: M = 3.71, SD = .98). For a graphical view, please see Figure 7.
Discussion

The results show that the acceptability of the technologies and methodology used during the workshop was high. This is a crucial first step for the incorporation of ICT tools in educational contexts (Sherman, & Howard, 2012). Regarding the workshop, the participants’ assessment and opinion have been favorable either for pedagogy and teaching methodology, aspects that facilitated the understanding of the tools, and the innovative nature of the contents of the workshop. Moreover, in the evaluation, the teachers also highlighted the vast possibilities of using these tools in the future, highlighting their usefulness in teaching. This is an important point, given that during the course, a significant feature used in the methodology was the training of the teachers in the selected technologies to use it in the future during their own practices with students. This suggests that it is likely that the acceptability is strongly related to how teachers value technologies for teaching and if they feel comfortable with them, as noted by other authors (Howard, 2011).

Another important point in our experience was that the positive appraisal of these ICT-based tools was independent from the e-learning experience, the background of the teachers, and the subjects that they teach. This can be interpreted as a good result in terms of the potential use of these technologies in the real context.

The participants also evaluated the technologies (PalMa, Eutopia, and e-Adventure) as being appropriate and easy to use. All systems received a score above 50% of the scale. However, one application stood out above the others: e-Adventure. These differences were evaluated as significant, according to the non-parametric Friedman test. Probably, given the characteristics of this system, the versatility and ability to adapt to different teaching contexts have been two of the features most appreciated by the participants. Nevertheless, compared to the other tools, this tool requires much training time. On the other hand, PalMa and Eutopia are more specific, and the variables manipulated by the users are smaller in number.

An important aspect to remark on is that all selected systems for the workshop have free access. This is a relevant standard of the workshop because it increases the possibilities of use these learning tools in the future.

The initiative of T3 Project involved facilitating the incorporation and distribution of new teaching tools in the teaching processes and in several learning areas. Our main goal was to influence student motivation through creative methodologies, taking advantage of their daily experience with technology. More specifically, the objective of the workshop was to show a series of tools designed to highlight the possibilities that the new technologies can offer, and to extend and disseminate the use of ICT in educational contexts. This means using more accessible teaching methodologies and making improvements that bring the university closer to the European Higher Education System.

Limitations of the study

There are some potential limitations to the interpretation and application of these results. The first limitation is that the entire sample was collected from the same university, and UJI is a pioneer in the field of these technologies. UJI was the first university in Spain to develop its own website and was a pioneer in the incorporation of technologies both in the classroom and in different studies developed in technological areas. This particular situation may suggest that teachers were more able and motivated to introduce technologies in their regular practice than teachers from other universities. Another limitation is that there is no way of measuring if the teachers made real use of the learning tools in their own practice once the workshop had finished. We do not know if using these new tools was useful to students’ practice nor do we know students’ opinion of the tools and their learning experience.

Future challenges

Regarding future challenges, our next step will be to implement the improvements in the ICT-based learning tools, following the opinion manifested by professors during the workshop. Also, we intend to evaluate the effect of this type of workshop in real practice and over the long term to see the effects of this practice on the satisfaction and motivation of both teachers and students. Besides, it would be interesting to replicate this work in other countries to
know the intercultural common or shared aspects and to do so in a cost-effective way, since the tools used are freely accessible and very easy to use.

Acknowledgements

This study has been funded in part by Ministerio de Ciencia e Innovación (PSI2010-17563), PROMETEO/2013/003 Fase II, ISIC/2012/012 Network of Excellence (Generalitat Valenciana, Conselleria de Educació), the CIBERobn (CB/06/03), and the Teaching to Teach with Technology Multilateral European Project (505169-LLP-1-2009-1-IT-KA3-KA3MP). We would like to thank the Educational Support Unit at UJI, where the workshop was carried out.

References


Bretón-López, J., Botella, C., Vizcaíno, Y., Quero, S., Baños, R., & Molés, M. (2010, October). El proyecto europeo T3 (enseñando a enseñar con nuevas tecnologías) [The European project T3 (teaching teaching with new technologies)]. Poster session presented at the VII Congreso Nacional de la Asociación Española de Psicología Clínica y Psicopatología, Benicàssim, Castellón de la Plana.


Wiki-Mediated Activities in Higher Education: Evidence-Based Analysis of Learning Effectiveness Across Three Studies

Panagiota Altanopoulou¹, Nikolaos Tselios¹*, Christos Katsanos²,³, Maria Georgoutsou¹ & Maria-Antonia Panagiotaki¹

¹ICT in Education Group, Dept. of Educational Sciences and Early Childhood Education, University of Patras, Patras, Greece // ²School of Science and Technology, Hellenic Open University, Greece // ³Department of Business Administration, Technological Educational Institute of Western Greece, Greece // galtanopoulou@gmail.com // nitse@ece.upatras.gr // ckatsanos@eap.gr // mariageorgoutsou@gmail.com // antoniapanag@gmail.com

*Corresponding author

(Submitted February 28, 2014; Revised December 30, 2014; Accepted March 8, 2015)

ABSTRACT

A considerable interest in using Web 2.0 technologies such as wikis in education has been observed recently. Despite the advantages of the wiki technology, a number of questions concerning design of appropriate activities and their learning effectiveness remain open. In this paper, we present the results of three activities involving first-year university students using wikis to learn basic concepts related to information technology. Three activities of different classes were designed and delivered to the students using the framework proposed by West and West (2009). In all studies, a one-group pretest–posttest design was adopted. Results illustrated significant improvement in learning outcomes, particularly for students with low initial performance. The average students’ questionnaire score jumped from 39.0/100 to 57.3/100. No significant effect of both students’ role in the activity and their school stream on learning gain was observed. Finally, regardless of the activity’s class and learning goal, a persistent pattern of high learning gain was observed.

Keywords

Web 2.0, Wikis, Learning activity design, Project-based learning, Collaborative learning

Introduction

The Internet and information and communication technologies (ICT) have a profound impact on our societies (Tapscott, 2009). As a result, the mode of learning is changing rapidly in a digital age (Palfrey & Gasser, 2008). During the recent years, there has been considerable interest in using Web 2.0 technologies in education.

Web 2.0 describes web-based technologies which emphasize on user-generated content. The content is created collaboratively with the opportunity to be shared with peers. Web 2.0 technologies constitute a fertile ground for building project-based learning activities (Duffy & Kirkley, 2004) and create real learning communities because students participate actively in the learning process (Pieri & Diamantini, 2014). In addition, engagement in Web 2.0-mediated activities seems to positively influence both active involvement as well as one’s motivation, which are recognized as key issues affecting student performance and learning (Benek-Rivera & Matthews, 2004; Cole, Feild, & Harris, 2004; Waycott, Bennett, Kennedy, Dalgarno, & Gray, 2010a; Waycott et al., 2010b). Most participants report positive experiences when they use Web 2.0 technologies (Ching & Hsu, 2011; Edirisingha, Rizzi, Nie, & Rothwell, 2007; Janossy, 2007). Furthermore, studies suggest that usage of specific Web 2.0 technologies can enhance student learning and collaboration (Carter, 2009; Cormode & Krishnamurthy, 2008; Greenhow, Robelia, & Hughes, 2009; Papastergiou, 2009; Selwyn, 2007). On a broader context, a recent meta-study (Means, Toyama, Murphy, Bakia, & Jones, 2009) concludes that students in online learning conditions performed modestly better than those receiving face-to-face instruction.

Despite the positive aspects of Web 2.0 technologies, a number of questions concerning design of appropriate activities and their learning effectiveness remain open (Gray, Thompson, Sheard, Clerehan, & Hamilton, 2010; Tselios, Daskalakis, & Papadopoulou, 2011; West & West, 2009). As it happens with other technologies used in education, there is an implicit perception that wikis could be instantaneously useful in the educational process without tackling the challenges related to their efficient integration in the educational context. For instance, Ching
and Hsu (2011) found that the activity goal needs to be shared in order to promote collaboration. Moreover, Ullrich et al. (2008) concluded that students using micro-blogging encouraged one another to participate and unconstrained active participation resulted in distractions.

Among the Web 2.0 technologies, wikis seem to offer the most dynamic collaboration possibilities (West & West, 2009). A wiki typically offers the ability to freely edit a website, providing features to add, modify, and delete pages as well as to integrate hypermedia (Leuf & Cunningham, 2001). The adopted interaction model is similar to that of a rich text editor with features of collaboration awareness, such as recent changes, actions carried out per participant, and edit-locking functionality (Tselios, Avouris, & Komis, 2008). This open nature of the wiki technology creates significant opportunities for learning (Mindel & Verma, 2006; Raman, Ryan, & Olfam, 2005; Wheeler & Wheeler, 2009). However, it may also be a major obstacle if the context and objectives of the activity are not well defined and/or not effectively communicated to the participants (Jones, 2007; Parker & Chao, 2007).

A wiki, by its very nature, facilitates quick content and organization deployment, which in turn increases the possibility of introducing inaccurate or incredible information or quoting unsubstantiated opinions. However, a wiki also enables all participants to edit and improve the provided content. This process of study, identification, and correction of content through reflection provides the opportunity for educational approaches compatible with socio-cultural views of learning (Cress & Kimmerle, 2008). Wikis as a collaboration tool help students to write better (Mak & Coniam, 2008), promote writing skills (Wheeler & Wheeler, 2009), and can support collaborative knowledge creation (Raman et al., 2005; Wagner, 2004).

Despite the encouraging results and the increasing adoption of wikis as a tool for collaborative learning, there is still the question of how to effectively integrate them into the educational process and which are, if any, the learning gains for the students involved into such an activity (Mason & Rennie, 2008; Pallof & Pratt, 2007). Currently, few studies provide rigorous, evidence-based results on the effectiveness of learning activities mediated by wikis (Biasutti & EL-Deghaidy, 2014; Hadjerrouit, 2014; Hazari, North, & Moreland, 2009; Heimbuch & Bodemer, 2014; Popescu, 2014; Salaber, 2014; Wheeler & Wheeler, 2009).

Preliminary results suggest that social, organizational, and cultural factors of the learning context are the important elements for effective use of wikis in educational practice and not the intermediary technology itself (Twu, 2009). Mindel and Verma (2006) also report that an empty wiki available to online students is not enough. In addition, Raman et al. (2005) indicate that wiki activities aren’t successful if there is not any proper planning and familiarity with this technology. Furthermore, Baltzersen (2010) found that students’ perceptions are positive if the wiki-based learning activity is carefully planned. As West and West (2009) state (p. 21): “Without context and support, online groups can experience unbalanced participation, a lack of progress and direction, mistrust, misunderstandings, and conflicts.” In such cases, the wiki technology may have little impact on student engagement (Cole, 2009; Ebner, Kickmeier-Rust, & Holzinger, 2008).

West and West (2009) have proposed a structured framework to guide the design of wiki-based activities. They identify the following critical factors to add context to the wiki environment: (a) establish a purpose for the wiki project, (b) define and classify the learning goals of the wiki project, (c) design a rich context and problem that support the achievement of the purpose and goals, (d) prepare students for work in the new environment, (e) promote a collaborative process through which active, social learning can take place (West & West, 2009, p. 22). Each team member should have a role with specific responsibilities, since students in group projects might not contribute the same amount of work (Elgort, Smith & Toland, 2008). In addition, according to this framework, wiki-based learning activities are grouped into three classes: knowledge construction, critical thinking, and contextual application.

However, there is a lack of rigorous studies demonstrating the learning effectiveness of such a structured framework. In addition, it remains unknown whether students’ role in the wiki activity affects their learning gain. Furthermore, students’ secondary education stream may also affect their learning gain in such activities because students of different streams might have a different manner of thinking. The aim of this paper is to investigate the effect of rigorously designed wiki-based activities, according to the framework proposed by West and West (2009), on the learning outcome. In specific, this paper investigates:

- students’ learning gain (if any) after each wiki-based activity
- whether the students with lower pretest score benefited from the activity at least to the same extent as students with higher pretest score
• whether students’ learning performance was affected by their secondary education stream
• whether students’ learning performance was affected by their assigned role

To this end, we designed and implemented three wiki-mediated activities, one for each class (West & West, 2009). In the first activity, the goal was to discuss implications of online social networking. The objective of the second activity was to learn basic aspects of search engines. In the third activity, students applied a usability evaluation technique. A one-group pretest–posttest experimental design was adopted.

The paper is organized as follows. Initially, the research method, the profile of the participants and the design of the activities are presented. Subsequently, the research results are presented, focusing on learning outcomes as assessed by appropriately designed pre-and posttest questionnaires. The paper concludes with a discussion on the findings and directions for future research.

Method

The main goal of this paper was to investigate possible differentiations of students’ learning gain (if any) across the three wiki-based activities. A one-group pretest–posttest design was adopted (Oncu & Cakir, 2011) to measure the learning gain as a result of students’ engagement in the wiki-mediated activity. Specifically, in order to examine the extent of knowledge and understanding before and after involvement in each activity, students responded to an appropriate online test comprising from 35 up to 40 multiple-choice questions with four answer options. The participants attended courses offered in the Department of Educational Sciences and Early Childhood Education (DESECE) at the University of Patras.

Participants

Twenty-four female first-year university students, aged 17–40 (M = 19.2, SD = 4.5) participated in the first activity. Students were attending a non-compulsory academic course entitled Introduction to Web Science. The activity took place from 26 May 2010 to 6 June 2010.

The second activity involved 146 students, 144 female, aged 17–40 (M = 19.2, SD = 3.6). Participants were attending a compulsory academic course entitled Introduction to ICT, and the activity took place from 29 November 2010 to 16 December 2010.

Thirty-six first-year university students, 35 female, aged 18–24 (M = 19.4, SD = 1.2), participated in the activity. Participants were attending a non-compulsory academic course entitled Introduction to Web Science, and the activity took place from 3 May 2011 to 20 May 2011.

Procedure

Participation in all three activities was one of the four compulsory mini-projects given to the students. The procedure was the following: First, an instruction on the wiki’s basic functionality was given to the students. Subsequently, the students familiarized with the wiki tool and practised upon representative tasks, such as creating and editing text, discussing and commenting aspects of the document and inserting and editing photos and videos. Next, a compulsory assignment was presented to them in the form of a wiki, constructed by the researchers. Each wiki included the objectives of the assignment, its structure, detailed implementation instructions, the expected learning outcome, the evaluation criteria, and the representative support material (West & West, 2009). The students worked remotely and, after the assignment deadline, they had to briefly present their work. The students were divided into groups comprising four or five members each. They were allowed to freely form their groups without any restrictions. Each team member was assigned a specific role in the group (West & West, 2009), such as Collector or Organizer. The available roles and associated responsibilities are delineated in the following. Each project was graded by the researchers on a scale from 1 to 100, according to the evaluation criteria. The score was multiplied by the number of the group members (4 or 5) and was given to the students, who were asked to distribute these points among themselves.
Research materials

For all activities, the Wikispaces service (www.wikispaces.com) was used both for the activity announcement and as the platform provided to the students to construct their wiki. The online questionnaire service Survey Monkey (www.surveymonkey.com) was used to create and distribute the questionnaires of the study. Afterwards, the obtained data were organized and analyzed using Excel 2010 and SPSS v20.0. The initial presentation of the activity to the students and the completion of the questionnaire took place at the departments’ computer lab.

The questionnaire was completed by the students at the beginning and the end of each learning activity. It included both demographic and knowledge-acquisition questions. The former were completed only before the wiki-mediated activity and collected personal information regarding Internet and wiki use and adoption. The latter were used as a student learning-assessment instrument and were primarily related to the Facebook service and the wiki tools (first activity), general information about Google’s history and services (second activity), and basic aspects and implications of usability evaluation in general and heuristic evaluation in particular (third activity). The students were not informed that they would be asked to complete the questionnaire at the beginning or the end of each activity.

Description of the activities

The design of each activity followed the structured framework proposed by West and West (2009). Students’ learning was expected to be achieved by their being engaging into four processes: (a) information seeking and retrieval, (b) argumentation development and refinement to support their thesis in the context of the assignment, (c) collaboration among members, and (d) their involvement with the wiki. Furthermore, in the third activity students had to evaluate a website, which was an additional learning objective for them. In all activities, the topic was selected because it was a notable part of the course’s overall outline.

In the first activity, available at http://ergastiriowiki.wikispaces.com, the learning goal was to understand and explore basic concepts of social networks, such as Facebook. Students were engaged in a debate exploring opposing sides of the Facebook service. On the one side, involving three out of the six groups with random selection, the students were responsible for directing Facebook’s public relations and had to defend against media allegations of Facebook being responsible for a variety of individual and social problems. On the other side, involving the rest groups, the students were asked to compose a critical essay about Facebook, which would focus on its negative impact on the users’ social lives, assuming that they were the editors of a newspaper with considerable prestige, social acceptance and validity. The wiki included four segment topics on which students relied to accomplish their task: (a) history and timeline of Facebook, (b) information about its rapid adoption, (c) reasons for Facebook’s popularity, and (d) arguments for/against Facebook.

In the second activity, available at http://googleactivity.wikispaces.com, the learning goal was to understand basic aspects and implications of search engines. The wiki included ten segment topics on which students relied in order to accomplish their task: (a) Google’s founders, (b) Google’s history, (c) the page-rank algorithm, (d) search techniques, (e) the technological infrastructure of the search engine, (f) Google’s working environment, (g) services provided, (h) Google’s business model, (i) usage of search engines in education, and (j) Google as a monopolist threat.

In the third activity, available at http://web-usability.wikispaces.com, the students had to study and present the most widely adopted usability evaluation techniques and include basic definitions about human computer interaction and user interface design. In addition, they had to delineate the process required to carry out a heuristic evaluation (Nielsen, 1994). Finally, they had to cooperatively evaluate the departments’ website (www.ecedu.upatras.gr), using heuristic rules (Nielsen, 1994).

The first activity required critical thinking, and students had to debate exploring opposing sides of the Facebook service. The second activity engaged students in knowledge construction processes. The third activity belongs to the third class of contextual application, in which students had to apply the heuristic rules that they had learned to evaluate their department’s website.
In all activities, the students had to create their own wiki, in which they would develop the theme of the assignment. An exemplary wiki was constructed by the researchers and provided to the students to support them in organizing their wiki and material and in structuring their arguments. Students had to cover the topics of the exemplary wiki, which were organized into subsections (Figure 1). For each topic and sub-topic, an indicative outline and specific arguments were given to the students to develop as well as supporting material and references, mainly in the form of hyperlinks (Figure 1).

![Figure 1. Screenshot of the second wiki-based activity presented to the participants.](image)

However, the context was not restrictive and the students were encouraged to use additional arguments and material. The use of additional material was not only desirable, but also a discrete evaluation criterion. Given that the open nature of Web 2.0 tools could lead to inappropriate use of content from other sources (Huijser, 2008; Waycott et al., 2010b; West & West, 2009), it was stressed that usage of other’s work should follow specific rules. The rest of the evaluation criteria were text relevance, text clarity, argument originality and reasoning, compliance to the provided structure and format guidelines, material appropriateness and richness, and appropriate use of references.

![Figure 2. Screenshot of a wiki developed by students](image)

All team members were writers and text editors, but each team member was assigned a role with specific responsibilities so that all students involved in a group would know what to expect from each other (West & West,
In all activities, there were four different roles: (a) Collector, responsible to obtain and organize appropriate material relevant to the subject undertaken by the team; (b) Organizer, responsible to examine the material collected by the Collector for consistency and relatedness with the objectives of the scenario; (c) Editor-in-Chief, responsible for composing the text with the basic arguments according to the objectives of the scenario and posting on the wiki the text and the material produced by the Collector and Organizer; (d) Controller, responsible to check the contents of the work in terms of its appropriateness, consistency, completeness, structure, and compliance with the objectives of the scenario. The students’ identity and role were published on the wiki used to describe the activity. Students were instructed to mainly carry out tasks that were related to their assigned role, but they could potentially carry out tasks that were not part of their role. However, the follow-up questionnaire confirmed that the students primarily focused on their initially assigned role. Figure 2 presents an example of a wiki produced by the students involved in the Facebook-related activity.

Results

All in all, we collected data from three wiki-mediated activities involving 206 university students. Table 1 presents participants’ demographic-related information in our dataset.

<table>
<thead>
<tr>
<th>Table 1. Participants’ demographic-related information in our dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity1: Facebook</td>
</tr>
<tr>
<td>Sample size</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>School stream</td>
</tr>
<tr>
<td>Theoretical</td>
</tr>
<tr>
<td>Technological/Scientific</td>
</tr>
<tr>
<td>Web-usage frequency</td>
</tr>
<tr>
<td>Almost none</td>
</tr>
<tr>
<td>Some times in a month</td>
</tr>
<tr>
<td>Some times in a week</td>
</tr>
<tr>
<td>Everyday</td>
</tr>
<tr>
<td>Prior wikis usage</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

First, reliability analysis was conducted for the three learning assessment questionnaires used. Reliability refers to the extent to which an instrument such as a questionnaire yields the same results under consistent conditions (Nunnally & Bernstein, 1994). It is most commonly measured using Cronbach’s alpha, which is a measure of internal consistency. The questionnaires used in both the first and third activities had good internal consistency; alpha = 0.77 and alpha = 0.86 respectively. The questionnaire used in the second activity did not have sufficient reliability (alpha = 0.65) to meet the typical minimum standard of 0.70 (Nunnally & Bernstein, 1994). Six questions increased the alpha to 0.72 when deleted and thus were excluded from subsequent analysis.

Following Nelson et al.‘s (2009) rationale, we used the normalized learning gain defined as:

\[
G = \frac{post_{score} - pre_{score}}{max_{score} - pre_{score}}
\]

This equation has the advantage of “normalizing the observed gain (the numerator) against the amount of possible learning that could be achieved (the denominator)” (Nelson et al., 2009, p. 1797), and thus allows for fair comparison of learning gains for students with different pretest scores.

Table 2 presents descriptive statistics of the collected dependent variables. Overall, students had similar initial scores in the knowledge assessment questionnaire and achieved higher scores after taking part in the wiki-mediated learning activity (Figure 3).
Table 2. Dependent variables grouped by wiki activity

<table>
<thead>
<tr>
<th>Activity</th>
<th>N</th>
<th>Pretest score [0–100] Mean ± 95% CI</th>
<th>Posttest score [0–100] Mean ± 95% CI</th>
<th>Normalized learning gain [%] Mean ± 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity 1: Facebook</td>
<td>24</td>
<td>37.7 ± 4.4</td>
<td>53.3 ± 6.1</td>
<td>22.4% ± 12.5</td>
</tr>
<tr>
<td>Activity 2: Google</td>
<td>146</td>
<td>40.2 ± 1.8</td>
<td>58.7 ± 2.3</td>
<td>29.5% ± 4.0</td>
</tr>
<tr>
<td>Activity 3: Heuristics</td>
<td>36</td>
<td>35.2 ± 3.7</td>
<td>54.1 ± 6.4</td>
<td>26.4% ± 11.9</td>
</tr>
<tr>
<td>All activities</td>
<td>206</td>
<td>39.0 ± 1.5</td>
<td>57.3 ± 2.1</td>
<td>28.2% ± 3.7</td>
</tr>
</tbody>
</table>

Figure 3. Students’ scores in the assessment questionnaire per wiki activity

An initial analysis was conducted to investigate whether there was an effect of activity (three levels) on students’ pretest score, posttest score, and learning gain. The assumption of normality was violated for all three dependent variables in at least one level of the independent variable (Shapiro-Wilk tests, p < 0.01). Levene’s test indicated that the assumption of homogeneity of variance was also violated for both the post-score (F(2,203) = 5.946, p < 0.01) and learning gain (F(2,203) = 4.884, p < 0.01). Thus, Kruskal-Wallis one-way ANOVA, a non-parametric test, was applied. Results showed no significant effect of activity on students’ pretest score (H(2) = 5.232, p = 0.073), posttest score (H(2) = 3.980, p = 0.137), and learning gain (H(2) = 1.004, p = 0.605). These findings support conducting analysis on the aggregated cross-activity dataset.

Additional analysis investigated prior experience of using wikis and web-usage frequency as potential covariates. Results showed no significant effects for both variables and thus they were excluded from subsequent analysis. It should be noted that our dataset is gender-skewed. Thus, any findings reported in the following might not be generalizable to male students involved in wiki-mediated learning activities. In all subsequent statistical analyses, effect sizes were calculated according to Field (2009).

Learning effectiveness of the wiki-mediated activities

A Wilcoxon signed rank test investigated differences between students’ pretest and posttest scores in the knowledge assessment questionnaire. A non-parametric test was selected since the distribution of the differences in the dependent variable (test score) between the two related conditions deviated significantly (D(206) = 0.984, p < 0.05) from a normal distribution. Results indicated that students achieved significantly higher (z = 10.698, p < 0.001, r = 0.527) test scores after participating in the wiki-mediated learning activity. This large effect size (Cohen, 1992) demonstrates the learning effectiveness of properly designed wiki-mediated learning activities in the context of higher education.

Additional analysis on a per-study basis demonstrated the same pattern. Wilcoxon signed rank tests after Bonferroni correction showed that students’ achieved significantly higher scores in the knowledge assessment questionnaire after participating in the first activity (z = 3.373, p < 0.001, r = 0.487), the second activity (z = 9.492, p < 0.001, r =
0.555), and the third activity \(z = 3.812, p < 0.001, r = 0.449\). Figure 3 presents students pretest and posttest scores per wiki activity.

### Learning gain for students with low and high pretest performance

Table 3 presents measured dependent variables grouped by students’ initial performance as assessed by their pretest scores. To this end, we recoded our dataset as in the following: students with pretest score below or equal to the median were assigned in the low pretest performance condition \((N = 110)\), whereas the rest were assigned in the high pretest performance condition \((N = 96)\). Posttest scores were similar for both students with low \((56.1/100)\) and high \((58.6/100)\) initial performance. However, students with low initial performance achieved higher learning gain \((36.0\%)\) compared to students with high initial performance \((19.2\%)\).

<table>
<thead>
<tr>
<th></th>
<th>Pretest score [0–100]</th>
<th>Posttest score [0–100]</th>
<th>Normalized learning gain [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± 95% CI</td>
<td>Mean ± 95% CI</td>
<td>Mean ± 95% CI</td>
</tr>
<tr>
<td>Low initial performance</td>
<td>30.9 ± 1.1</td>
<td>56.1 ± 2.7</td>
<td>36.0% ± 3.9%</td>
</tr>
<tr>
<td>High initial performance</td>
<td>48.3 ± 1.5</td>
<td>58.6 ± 3.2</td>
<td>19.2% ± 6.3%</td>
</tr>
</tbody>
</table>

A two-tailed Man-Whitney U test investigated the effect of students’ initial performance on their normalized learning gain. A non-parametric test was selected because both the assumptions of normality (Shapiro-Wilk tests, \(p < 0.001\)) and homogeneity of variance (Levene’s test, \(F(1,204) = 5.696, p < 0.01\)) were violated. It was found that students with low initial performance achieved a significantly higher \((z = 4.108, p < 0.001, r = 0.286)\) learning gain compared to students with high initial performance. Considering the whole dataset, a significant negative correlation \((r_s = –0.293, p < 0.001)\) was also found between students’ pretest score and normalized learning gain.

These findings suggest that wiki-mediated learning activities might be more beneficial to students with lower initial performance. However, more investigation and additional studies are required to both verify this finding and identify the specific reasons for this pattern of learning gain.

### Effect of students’ school stream on learning gain

The students’ performance according to the stream that they choose in the secondary education’s curriculum is presented in Table 4. The pretest mean score of students who attended the theoretical stream was slightly lower \((38.7/100)\) compared to that of students of the technological or the scientific stream \((39.9/100)\). Students coming from a theoretical background in secondary education also achieved lower posttest scores \((56.6/100)\) compared to those with a technological or scientific background \((59.4/100)\).

<table>
<thead>
<tr>
<th></th>
<th>Pretest score [0–100]</th>
<th>Posttest score [0–100]</th>
<th>Normalized learning gain [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± 95% CI</td>
<td>Mean ± 95% CI</td>
<td>Mean ± 95% CI</td>
</tr>
<tr>
<td>Theoretical</td>
<td>38.7 ± 1.6</td>
<td>56.6 ± 2.4</td>
<td>27.2% ± 4.4%</td>
</tr>
<tr>
<td>Technological/Scientific</td>
<td>39.9 ± 3.7</td>
<td>59.4 ± 4.2</td>
<td>30.9% ± 7.4%</td>
</tr>
</tbody>
</table>

A two-tailed Man-Whitney U test investigated the effect of students’ school stream on their normalized learning gain. A non-parametric test was selected since the assumption of normality was violated for the group of students who attended the theoretical school stream \((D(154) = 0.918, p < 0.001)\). Results showed no significant difference \((z = 0.731, p = 0.465)\) in the learning gain achieved by students who attended the theoretical curriculum and who attended the technological or scientific curriculum.

This finding suggests that wiki-mediated learning activities are beneficial to all students, regardless of their secondary education stream of studies. However, one should be cautious with this finding, given that approximately 75% of the students in our sample came from a theoretical school stream.
Effect of students’ role in the activity on their performance

Table 5 presents students’ performance according to their assigned role in the wiki-based learning activity: Collector \((N = 76)\), Organizer \((N = 40)\), Editor-In-Chief \((N = 46)\), and Controller \((N = 44)\).

<table>
<thead>
<tr>
<th>Role</th>
<th>(N)</th>
<th>Pretest score ([0–100]) Mean ± 95% CI</th>
<th>Posttest score ([0–100]) Mean ± 95% CI</th>
<th>Normalized learning gain [%] Mean ± 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collector</td>
<td>76</td>
<td>39.9 ± 2.6</td>
<td>59.7 ± 2.9</td>
<td>30.9% ± 5.7%</td>
</tr>
<tr>
<td>Organizer</td>
<td>40</td>
<td>39.7 ± 3.9</td>
<td>58.3 ± 5.1</td>
<td>28.2% ± 9.6%</td>
</tr>
<tr>
<td>Editor-in-Chief</td>
<td>46</td>
<td>38.8 ± 3.1</td>
<td>53.5 ± 5.2</td>
<td>22.7% ± 8.9%</td>
</tr>
<tr>
<td>Controller</td>
<td>44</td>
<td>37.1 ± 2.8</td>
<td>56.2 ± 4.5</td>
<td>29.1% ± 7.8%</td>
</tr>
</tbody>
</table>

A Kruskal-Wallis one-way ANOVA test investigated the effect of students’ role on their normalized learning gain. A non-parametric test was selected since the assumption of normality was violated for the students with the role of Collector \((D(76) = 0.885, \ p < 0.001)\), Editor-in-Chief \((D(46) = 0.949, \ p < 0.05)\) and Controller \((D(44) = 0.936, \ p < 0.05)\). Results showed no significant effect of students’ role on their learning gain \((H(3) = 2.674, \ p = 0.445)\).

This finding suggests not only that wiki-mediated learning activities were beneficial to all students, regardless of their specific role, but also that the responsibilities of each role were well-distributed.

Conclusions and discussion

In this paper, the results of three activities investigating the effectiveness of wiki-mediated learning were presented. Two hundred and six first-year university students participated in the activities. The design of the activities followed the framework proposed by West and West (2009). Evaluation of the learning effectiveness of the activities was carried out using a one-group pretest–posttest design. The results showed significant improvement in learning outcomes, particularly for students with low initial performance. The average students’ questionnaire score jumped from 39.0/100 to 57.3/100. The students with low initial performance (below or equal to the median initial score) showed an improvement of 25.2 percentage points, whereas the students with high initial performance (above median initial score) showed an improvement of 10.3 percentage points. In addition, in all three activities a comparable and significant learning gain was observed. The persistence of the results strongly indicates that students learn with wikis regardless of the activity’s class and subject.

In addition, no significant variation between the students’ secondary education curriculum (i.e., school stream) and their learning outcome in the wiki-mediated activity was observed. Given that all the three activities were both mediated and related to information technology, students coming from a technological or scientific background were expected to achieve higher learning gain. However, it was found that wiki-mediated learning activities were beneficial to all students, regardless of their previous knowledge in secondary education.

Furthermore, it was found that learning gain is not related with the student’s role in the activity. This is in line with previous research (Strijbos, Martens, Jochems, & Broers, 2004; Tselios, Altanopoulou, & Katsanos, 2011; Tselios, Altanopoulou, & Komis, 2011). However, the introduction of roles can help students who work collaboratively to build knowledge (Schellens, Van Keer, De Wever, & Valcke, 2007) and can increase cohesion, responsibility, and awareness in group members (Strijbos et al., 2004). By contrast, De Wever, Keer, Schellens, and Valcke (2010) noticed a varying impact of roles on knowledge construction.

All in all, this paper makes the following contributions related to the effectiveness of wiki-based learning activities:
- Significant and persistent learning gain was found across three activities, which are of different class according to the framework proposed by West and West (2009).
- The learning gain was found to be significantly higher for students of low initial achievement compared to students with high initial achievement.
- Students’ role in the activity did not significantly affect their learning gain.
• Student’s school stream did not significantly affect their learning gain.

These findings provide evidence that a wiki-based activity with a suitable context and support can substantially facilitate students to achieve higher levels of learning. Given that there is a lack of rigorous studies demonstrating the learning effectiveness of the framework proposed by West and West (2009), these are important findings for teachers, university instructors, instructional designers, and even technologists who are developing wikis and wiki-like platforms.

However, this research is not without limitations. The obtained results do not explain how the students have benefited from their involvement in the wiki activity. Moreover, it is not known to what extent the students were improved in other non-cognitive aspects, such as self-organization, collaboration, attitudes toward technology, and openness (Tapscott, 2009), which are considered important for completing a wiki project. Further studies are required to investigate these issues. Future research goals also involve the design and deployment of additional wiki-based activities in a variety of educational settings (both in tertiary as well as in secondary education) and the investigation of learners’ behavioral intention to use wiki technology using technology acceptance models (Tselios, Daskalakis, & Papadopoulou, 2011). In addition, investigation of the interaction between students’ observed activity in a wiki and the learning outcome (Katsanos, Tselios & Avouris, 2010; Tselios & Avouris, 2003) will be also examined. Finally, a comparison with other Web 2.0 technology-mediated activities also constitutes a future research goal.

References


In “Computer Games for Learning,” Richard E. Mayer presents a strong platform for understanding the current state of research in educational games and gamification. Unlike many books focusing on games or gamification, Mayer rejects any statements on educational games which are not supported by well-constructed research. The focus of this book is on that research and not the topic of games or gamification itself per se. It presents the concepts and methodologies of conducting research within the field of educational games, along with a background of work that has already been conducted, focusing on a series of studies he and his team have conducted to investigate their significance.

Researchers interested in studying gamification and the use of games within learning spaces will find this book a useful, thorough and, most of all, a pragmatic addition to their library. The introduction of “Computer Games for Learning” explores the issues which researchers will face, including a description of the domain of educational games. The three chapters that make up the Introduction provide an excellent platform for the rest of the book, distinguishing both the intention of the work and the method that Mayer uses.

The foundation Mayer creates is used both in the presentation of the work and in the studies conducted by his team. Other researchers will find this foundation useful for their own research, as something of a test to ensure that their studies are properly focused and removed from the hype and bias so often found within gamification literature. It is composed of three parts, presented as “Value-Added Designs,” “Cognitive Consequence Designs,” and “Media Comparison Designs.” These three inquiry categories are better described as “Determining the effectiveness of adding a feature to a game,” “Determining the educational value of an off-the-shelf game,” and “Determining the value of an educational game versus some conventional educational context”. These, Mayer suggests, are the areas in which research can be most fruitful and in which the results of research can be most sure.

The body of the work, Part II (Chapters 4 to 7) of “Computer Games for Learning,” presents the statements from gamification advocates, findings of other researchers, and the findings of his own team. Here, Mayer focuses on the propositions made by the community of educational game researchers and how those propositions stand against the testing methodologies proposed in the introduction of the book. Researchers interested in gamification will find the studies presented to be well-constructed and replicable, with methods that can be re-used for other purposes and with conclusions that inspire new questions and further research.

This section is also of interest to those interested who are outside of the research community, such as educators or educational administrators. Findings are presented in a clear and unambiguous manner and cover a wide array of subjects within the field of gamification. More generally, the book as a whole presents a clear methodology for separating fact from fiction which is approachable for the layperson developing their own gamification programs. For such a reader, Mayer presents each question in the same format, including presentation of each individual problem and how that problem was approached before presenting and discussing findings. Mayer does not shy away from pointing out where findings contrast conventional wisdom, instead drawing attention to those factors which are in conflict. For example, while concepts such as using voiced instead of text narration are strongly supported by
testing, others are assumed to be valid are found wanting – the idea of providing both text and voiced narration simultaneously is challenged, along with the inclusion of competition in a learning game.

While subtle, the underlying theme of the book is a useful one for researchers, enthusiasts and laypersons to keep in mind. This theme is stated most clearly in a discussion of a “Media Comparison,” which indicates that the support for educational games as useful pedagogical tools is not as strong as is popularly believed, with computer-based slideshows being more educational in two out of the three principles examined. Throughout the book, Mayer gently suggests that enthusiasm has overtaken critical thinking in the field of gamification. This book is an effort to rein in that enthusiasm, especially within the research community examining games for learning.

Unfortunately, this effort to rein in the enthusiasm in the gamification field is one of the major flaws of the book. Gamification as a field of research is complex, poorly structured, and dense. Despite the problems it introduces, the enthusiasm that is present in the field is an overall strength, generating diverse new directions of research. Without that energy, the book can be a little dry and difficult to go through. It is arguable that this is a necessary change in the field for it to be considered a “mature” field of research, but it may come as a shock to an enthusiastic researcher interested and deeply invested in the work. This still remains a valuable book for these individuals, though, perhaps more so. A dose of sober reflection on a field is often more valuable when that field is exciting and generating a lot of interest.

In conclusion, the book provides a clear, honest and critical perspective of research in educational game design. The appraisal of Dr. Jan L. Plass, as found on the back jacket is fitting: “Computer Games for Learning is an indispensable read for anyone interested in the field of games and learning.” While at times dense and technical, Mayer provides a much-needed analytical framework, providing tools for sober, rational, and realistic investigation into educational games and gamification.
Embracing Social Media: A Practical Guide to Manage Risk and Leverage Opportunity
(Book Review)

Reviewer:
Jon Dron
Athabasca University, Canada
jond@athabascau.ca

Book Details:
Embracing Social Media: A Practical Guide to Manage Risk and Leverage Opportunity
Written by Kristin Magette
2014, 116 Pages, Published by Rowman & Littlefield Publishers
ISBN 978-1-475-81328-9 (Hardback) 978-1-475-81329-6 (Paperback)

Audience
This slim volume is primarily aimed at administrators in American school systems that wish to actively implement a social media policy in order to make greater use of social media in their practice. It provides a formula gleaned from the experience of the author and her colleagues of an approach used in the Eudora School District in Kansas, US, to embed social media in the daily practices of district schools.

Structure and content
As the title suggests, the book is meant as a practical guide. It is clearly structured along the lines of a “how-to” manual. Each of the ten chapters deals with different, albeit overlapping, aspects of using social media in and for schools, and each concludes with summaries of recommendations that condense the contents of the chapter into easily digestible bullet points for action. The first chapter presents cases for embracing social media, most of which are variations on a theme of “the train has already left the station.” The second explores some dangers and provides arguments to use to persuade those who are reluctant to embrace social media. The third, fourth and fifth chapters talk, in turn, about the process of getting started, establishing policies, and designing processes and procedures. Chapter Six describes an approach to professional development for social media, mainly relying on lectures and workshops. Chapter Seven provides an assortment of advice on “best” practices, including ways to use some specific tools. Chapter Eight addresses teaching and learning, though is mostly concerned with the administrative side of this function. The last two chapters are concerned with communication (notably focusing on dissemination of information) and building communities, including a final chapter on dealing with “problem” posts, both from those outside the administration and those within the administration that need to disseminate sensitive information.

The good
Magette appears to have plenty of experience of implementing a social media policy in an American school district. There is a lot of specific advice born from experience, much of which makes good sense. Chapter Two, in particular, contains some sensitive arguments against prejudices and concerns that that may be encountered when trying to pull recalcitrant communities on board, and Chapter Seven provides a useful smorgasbord of suggestions about good practice in a few social media sites. In general, it is packed with recommendations, suggestions and patterns to follow that school administrators might find useful in implementing a social media policy.
The less good

The biggest problem with this book is that it turns a description of what was done in the Eudora School District into a prescription for what should be done elsewhere, without ever discussing why or whether it might be a good idea. Little evidence is provided as to whether the process followed here was a particularly effective. In fact, there is hardly anything to describe the effects that it had in the Eudora School District at all, though much information is provided about the process that was followed. What little we do get relating to the experience is mostly in the form of a few quoted fragments from participants that, though useful in revealing at least a little of the human side of it and some of the potential benefits, do not reveal any of the complexity involved, problems encountered, nor arguments and issues that emerged. There is almost no discussion of alternatives, different points of view or wicked problems that must be addressed. There is almost no reference to any previous literature or other interventions of a similar kind elsewhere. There is not even an appendix or list of references that might be consulted by those seeking more background and depth.

Apart from concerns that not all of the methods suggested might be good ones, the trouble with this approach is that there are probably no places in the World that are exactly like the Eudora School District. The content is highly parochial: it relates strongly to structures, laws, rules, norms and processes that are peculiar to the US public school system, with no consideration of the broader international context nor the fact that readers might not know how such things are organized and managed in America. With that in mind, it would have been helpful to have at least provided appendices to explain the terminology for those unfamiliar with this system because there are generic issues that would be relevant beyond this context.

There are enough tips, hints, implied stories, methods and techniques in this to pull something of value out of it if you already know what you are doing with social media, and perhaps have some theoretical or practical background in technology acceptance models and change management. However, this book is not made for experts with such knowledge. There is no theoretical or empirical support for anything prescribed in the book, no compelling arguments, no consideration of different approaches, no references for further reading, no examples to reflect upon, no weighing of issues, no exploration of the reasons behind decisions and barely any discussion of whether or how it worked. On those odd occasions that reasons are given they are almost always in relation to the context and needs of the Eudora School District and are seldom if ever backed up by research. This is a recipe book, pure and simple, for reproducing one pattern. A good number of the recipes are useful, well considered and sound. Unfortunately, it is not likely to be easy for the intended audience to distinguish these from others that are less sound, because there is no foundation, no reasoning, not even an appeal to authority to help come to an informed decision.

There are reasons to be concerned about some of the methods used here. The general attitude towards social media appears to owe more to ways commercial companies manage their social media presence than to exploiting the potential for creating rich, caring, participative networks of the sort we might hope to see in and between schools and their communities. The focus is largely on relationships between individuals in the community and the distinct corporate entity of the school itself, more than between the people within it. It turns people into roles, not human beings, and it overlays a hierarchical culture of control over what is natively a networked and bottom-up environment. This is more suited to marketing of corporations than networking of schools. There is much more in this about managing perceptions and disseminating controlled information than there is about the value of social media to support knowledge building and community.

Furthermore, the book promotes a very controlling and pedagogically under-informed attitude both to development and to community. For example, it is telling that training in the Eudora district began with lectures informing groups of 50-150 teachers of the district’s social media policies and procedures. Whether or not one approves of the intent, it seems very odd that social media were not the first and most central learning media used for this. An emblematic quote that summarizes this pervasive attitude of control and centralized management is that “leaders must hold up what we value, and moderate what we don’t value.” Such attitudes make sense in a closely controlled, role-based and hierarchically structured online community context, but they are not really appropriate in a networked social media context. At the very least, discussion and reflection are needed, as well as some consideration of alternative ways of seeing the issues.

Children are notable by their absence throughout the book, save as ciphers or subjects to administer. This is symptomatic of the general problem that the book is almost exclusively concerned with an administrator’s
perspective on how rather than why things should be done. The single chapter on teaching and learning (the order is telling) actually contains hardly anything about teaching apart from some second-hand reporting of shallow repurposing of classroom pedagogies to force children to use social media instead of exercise books, and it contains nothing about learning at all. No understanding or knowledge is shown of the distinctive pedagogies, learning opportunities and social forms of social media about which much has been written over the past decade or two. This cries out for more research support and more background than Magette provides. Again, even a reference or two for further reading would help.

Apart from the lack of reflective discussion, research, or analysis that I have commented on already, even on its own terms as an administrator’s handbook, there are some surprising omissions. One of the most notable of these is the problem of evaluation. Apart from a tiny section in Chapter Seven on basic analytics (that shows a worryingly limited understanding of the limitations of the technologies), there is hardly any mention of ways to discover the success or otherwise of the strategies and methods presented here. Several recommendations scattered throughout do signal that evaluation is very important but, in contrast to the detailed recipes given in the bulk of the book, very little advice is given on how to do it.

A related blind-spot is that there is little useful advice on change management, despite frequent acknowledgement that change is the one constant in social media. This is a particular problem because the book is heavily biased towards “free” social media platforms like Facebook and Twitter, over which members have no control. Such platforms frequently change their terms and conditions, pricing, technological features, and ownership. Because they are accountable to shareholders, not members, these changes can be highly disruptive or harmful, especially in the context of schools that have a duty of care and responsibility to children and their families. Unfortunately, almost all business models for such platforms rely on single-mindedly locking their members. Simply adding content to a platform, investing in creating pages and profiles, and building a social network, is usually more than enough to make backing out from the platform a difficult decision, even for individuals, all the more so for schools. This book has no consideration of an exit strategy, nor even of more than surface approaches to adapting to changes when they occur. One of the problems with the book’s prescriptive and non-discursive approach is that it does not provide the cognitive tools needed to think about such changes.

Following from that, especially given the culture of control that informs most of the book, it is puzzling that almost no consideration is given to the potential for using school- or district--controlled social media (e.g., those written using BuddyPress, Drupal or Elgg) that can be optionally linked with and published to other commercial social media accounts where desired, as well as to be fed from them. Many institutions have taken a home-grown path like this in order to preserve control and to support the needs for privacy and ownership of their stakeholders, keeping commercial social media close, but not so close that they dominate or are necessary to sustain engagement. Indeed, in many countries, privacy legislation would require such a model, regardless of pragmatic benefits. Relatedly, there is little serious consideration given to the needs of those who, for whatever reason, do not wish to be seen or heard on particular social media or that cannot easily access such a system. There are enormous risks of exclusion of those unwilling or unable to be farmed by large commercial social media sites, or those that wish to have more control of their own privacy, or those that do not have much access to online tools in the first place. A single short section on privacy in Chapter Six is concerned almost solely with meeting legislative requirements, not with addressing the complex challenges of identity and community that arise in such contexts.

Finally, the choice of social media examples is very limited, with Facebook and Twitter dominating, and a little lip service to other platforms like Instagram, YouTube and Vine. There are many different forms of social media of which these are a tiny sample, most of which get little or no mention. It is surprising that there is not more discussion of Wikipedia, the Khan Academy, LinkedIn, Ning, Pinterest (better still Learnist), StackExchange, Flickr, Wordpress or Google Plus, to mention but a few of the more obvious candidates that might be significant to schools. There are great differences as well as common features in these tools. At the very least, it would have been helpful to provide some guidance to the presumed novice audience, even if only as an appendix.

**Is this book for you?**

There is a limited market for this little book. It might be a useful recipe book to help support your first tentative steps into embracing social media if:
• You are an administrator in an American public school system, ideally in Kansas;
• You feel you should be doing something to address the changing landscape brought about by the growth of social media or are feeling left behind;
• You need a hand-holding step-by-step approach and do not wish to think about it too deeply;
• You hold a conventional set of attitudes to the roles of teachers and schools;
• You have some enthusiasm for but not much experience of social media.

In summary

It is, perhaps, a little unfair to review a book that makes no pretence to containing any research at all in a journal that is devoted to research, and it important to remember that most of the readers of this review are not the book’s target audience. However, with a bit more reference to previous work and to theory, and with a lot more analysis of the issues and complexities, this might have been a much more useful book for everyone. Even if it had a usable appendix of further reading, or some reference to examples that could be explored, it would have helped fill in some of the gaps that most of the readership would be likely to discover as soon as they attempted to follow this process. This book allows others to replicate the processes and patterns of the Eudora School Board but offers little support for those that wish to understand why or whether they might wish to do so.