Journal of
Educational Technology & Society

Published by
International Forum of Educational Technology & Society

July 2012
Volume 15 Number 3
Educational Technology & Society
An International Journal

Aims and Scope

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

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

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

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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

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

Editors

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

Editors’ Advisors

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

Editorial Assistant

Barbara Adamski, Athabasca University, Canada.

Associate editors

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

Assistant Editors

Yuan-Hsuan (Karen) Lee, National Chiao Tung University, Taiwan; Weichang Fang, National Sun Yat-sen University, Taiwan.

Advisory board

Igancio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-Ea Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J. R. Isaac, National Institute of Information Technology, India; Mohamed Jenni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasia Karoulis, Hellenic Open University, Greece; Paul Kirscheiner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Roh 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; Riichiro Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan.

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

## Special Issue Articles

- **Guest Editorial - Learning and Knowledge Analytics**  
  *George Siemens and Dragan Gasevic*  
  1-2
- **Social Learning Analytics**  
  *Simon Buckingham Shum and Rebecca Ferguson*  
  3-26
- **Integrating Data Mining in Program Evaluation of K-12 Online Education**  
  *Jui-Long Hung, Yu-Chang Hsu and Kerry Rice*  
  27-41
- **Translating Learning into Numbers: A Generic Framework for Learning Analytics**  
  *Wolfgang Greller and Hendrik Drachsler*  
  42-57
- **Design and Implementation of a Learning Analytics Toolkit for Teachers**  
  *Anna Lea Dyckhoff, Dennis Zielke, Mareike Bültmann, Mohamed Amine Chatti and Ulrik Schroeder*  
  58-76
- **Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade**  
  *M’hammed Abdous, Wu He and Chering-Jyh Yen*  
  77-88
- **A Multidimensional Analysis Tool for Visualizing Online Interactions**  
  *Minjeong Kim and Eunchul Lee*  
  89-102
- **Teaching Analytics: A Clustering and Triangulation Study of Digital Library User Data**  
  *Beijie Xu and Mimi Recker*  
  103-115
- **Analyzing Interactions by an IIS-Map-Based Method in Face-to-Face Collaborative Learning: An Empirical Study**  
  *Langin Zheng, Kaicheng Yang and Ronghui Huang*  
  116-132
- **Dataset-Driven Research to Support Learning and Knowledge Analytics**  
  *Katrien Verbert, Nikos Manouselis, Hendrik Drachsler and Erik Duval*  
  133-148
- **Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan**  
  *Leah P. Macfadyen and Shane Dawson*  
  149-163

## Full Length Articles

- **Investigating the Development of Work-oriented Groups in an e-Learning Environment**  
  *Chia-Ping Yu and Feng-Yang Kuo*  
  164-176
- **Alignment of Teacher and Student Perceptions on the Continued use of Business Simulation Games**  
  *Yu-Hui Tao, Chieh-Jen Cheng and Szu-Yuan Sun*  
  177-189
- **An Analysis of Students’ Academic Performance When Integrating DVD Technology in Geography Teaching and Learning**  
  *C. P. (Christo) Van der Westhuizen, Carisma Nel and Barry W. Richter*  
  190-201
- **Investigating Learner Affective Performance in Web-based Learning by using Entrepreneurship as a Metaphor**  
  *Ming-Chou Liu and Ming-Hsiao Chi*  
  202-213
- **A User-Centric Adaptive Learning System for E-Learning 2.0**  
  *Shiu-Li Huang and Jung-Hung Shiu*  
  214-225
- **Social Networks-based Adaptive Pairing Strategy for Cooperative Learning**  
  *Po-Jen Chuang, Ming-Chao Chiang, Chu-Sing Yang and Chun-Wei Tsai*  
  226-239
- **Exploring the Factors Influencing Learning Effectiveness in Digital Game-based Learning**  
  *Fu-Hsing Tsai, Kuang-Chao Yu and Hsien-Sheng Hsiao*  
  240-250
- **The Impact of Adapting Content for Students with Individual Differences**  
  *Raymond Flores, Fatih Ari, Fetli A. Inan and Ismahan Arslan-Ari*  
  251-261
An Ecological Approach to Learning Dynamics  
*Peeter Normak, Kai Pata and Mauri Kaipainen*

KnowledgePuzzle: A Browsing Tool to Adapt the Web Navigation Process to the Learner’s Mental Model  
*Iyad AlAgha*

Digital Competition Game to Improve Programming Skills  
*Julián Moreno*

A Comparison of Demonstration and Tutorials in Photo Editing Instruction  
*Cenk Akbiyik*

Student Engagement in Blended Learning Environments with Lecture-Based and Problem-Based Instructional Approaches  
*Omer Delialioğlu*

Invited article(s)

Development Patterns of Scientific Communities in Technology Enhanced Learning  
*Manh Cuong Pham, Michael Derntl and Ralf Klamma*

Book review(s)

Flexible Pedagogy, Flexible Practice: Notes from the Trenches of Distance Education (Editors: Elizabeth Burge, Chere Campbell Gibson and Terry Gibson)  
*Reviewer: Dermod Madden*
Guest Editorial - Learning and Knowledge Analytics

George Siemens and Dragan Gasevic
Athabasca University, Canada // gsiemens@athabascau.ca // dragang@athabascau.ca

The early stages of the internet and world wide web drew attention to the communication and connective capacities of global networks. The ability to collaborate and interact with colleagues from around the world provided academics with new models of teaching and learning. Today, online education is a fast growing segment of the education sector. A side effect, to date not well explored, of digital learning is the collection of data and analytics in order to understand and inform teaching and learning. As learners engage in online or mobile learning, data trails are created. These data trails indicate social networks, learning dispositions, and how different learners come to understand core course concepts. Aggregate and large-scale data can also provide predictive value about the types of learning patterns and activity that might indicate risk of failure or drop out.

The Society for Learning Analytics Research defines learning analytics as the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (http://www.solaresearch.org/mission/about/). As numerous papers in this issue reference, data analytics has drawn the attention of academics and academic leaders. High expectations exist for learning analytics to provide new insights into educational practices and ways to improve teaching, learning, and decision-making. The appropriateness of these expectations is the subject of researchers in the young but rapidly growing learning analytics field.

Learning analytics currently sits at a crossroads between technical and social learning theory fields. On the one hand, the algorithms that form recommender systems, personalization models, and network analysis require deep technical expertise. The impact of these algorithms, however, is felt in the social system of learning. As a consequence, researchers in learning analytics have devoted significant attention to bridging these gaps and bringing these communities in contact with each other through conversations and conferences. The LAK12 conference in Vancouver, for example, included invited panels and presentations from the educational data mining community. The SoLAR steering committee also includes representation from the International Educational Data mining Society (http://www.educationaldatamining.org).

This issue reflects the rapid maturation of learning analytics as a domain of research. The papers in this issue indicate LA as a field with potential for improving teaching and learning. Less clear, currently, is the long-term trajectory of LA as a discipline. LA borrows from numerous fields including computer science, sociology, learning sciences, machine learning, statistics, and “big data”. Coalescing as a field will require leadership, openness, collaboration, and a willingness for researchers to approach learning analytics as a holistic process that includes both technical and social domains.

This issue includes ten articles:

Buckingham Shum and Fergusson describe social learning analytics (SLA) as a subset of learning analytics. SLA is concerned with the process of learning, instead of heavily favoring summative assessment. SLA emphasizes that “new skills and ideas are not solely individual achievements, but are developed, carried forward, and passed on through interaction and collaboration”. As a consequence, analytics in social systems must account for connected and distributed interaction activity.

Hung, Hsu, and Rice explore the role of data mining in K-12 online education program reviews, providing educators with institutional decision-making support, in addition to identifying the characteristics of successful and at-risk students.

Greller and Drachsler propose a generic framework for learning analytics, intended to serve as a guide in setting up LA services within an educational institution. In particular, they emphasize the challenges of the soft dimensions of learning analytics such as ethics and the need for educators to develop competence (literacies) in interacting with data.
Dyckhoff et al. detail eLAT (exploratory learning analytics toolkit). eLAT is intended to give educators access to tools for visualizing teaching and learning activity with a primary benefit being the ability of teachers to self-reflect.

Abdous, He, and Yen discuss the results of hybrid analysis (educational data mining and regression analysis) in order to analyze student’s activity in live video sessions and their course performance. They conclude that educational data mining can convert “untapped LMS and EPR data into critical decision-making information which has the capability of enhancing students’ learning experiences”.

Kim and Lee suggest that the prominent analytics techniques function in isolation and, as a consequence, are one-dimensional. In response, they propose the Multidimensional Interaction Analysis Tool (MIAT). Multidimensional analysis can provide “more in-depth information about the learning process and the structure of online interactions”.

Xu and Recker share the results of a clustering study on how educators use a digital library tool called Instructional Architect. Their findings indicate three clusters of educators - key brokers, insular classroom practitioners, and inactive islanders – and suggest that analytics can be used to “predict which kinds of teachers are more likely to adapt technology tools such as digital libraries, and more importantly, how to help teachers become more effective digital libraries users”.

Zheng, Yen, and Huang evaluate the role of analytics in understanding information flows. Their Interactional Information Set (IIS) model seeks to explain the collaborative process and information activation that occurs through interaction between learners.

Verbert et al. identify the challenges that researchers face with regards to the availability of open data sets. These data sets are important in order for researchers to test new algorithms and compare results with the results of other researchers. To address this challenge, Verbert et al. present a framework for analyzing educational data sets and present future challenges around collection and sharing of data sets.

Macfadyen and Dawson consider resistance to institutional adoption of learning analytics from the perspective of change management theories, arguing “research must also delve into the socio-technical sphere to ensure that learning analytics data are presented to those involved in strategic institutional planning in ways that have the power to motivate organizational adoption and cultural change”. As learning institutions begin to deploy learning analytics, careful consideration of resistance factors can help to increase successful outcomes of enterprise-level analytics strategies.

During the Learning Analytics and Knowledge 2012 conference in Vancouver, a keynote speaker – Barry Wellman – described his experiences in the early 1970’s in helping to establish the field of social network analysis. Wellman stated that the activity and energy that he felt within the learning analytics field were comparable to those within social network analysis several decades ago. In putting together this special issue, we hope to provide a small, but meaningful, contribution to the growing numbers of researchers and academics who are turning their attention to data and analytics as a means to become better teachers and help learners become better learners.
Social Learning Analytics

Simon Buckingham Shum1* and Rebecca Ferguson2
1Knowledge Media Institute & 2Institute of Educational Technology // The Open University, Milton Keynes, MK7 6AA, United Kingdom // S.Buckingham.Shum@open.ac.uk // R.M.Ferguson@open.ac.uk
*Correspondence author

ABSTRACT
We propose that the design and implementation of effective Social Learning Analytics (SLA) present significant challenges and opportunities for both research and enterprise, in three important respects. The first is that the learning landscape is extraordinarily turbulent at present, in no small part due to technological drivers. Online social learning is emerging as a significant phenomenon for a variety of reasons, which we review, in order to motivate the concept of social learning. The second challenge is to identify different types of SLA and their associated technologies and uses. We discuss five categories of analytic in relation to online social learning; these analytics are either inherently social or can be socialised. This sets the scene for a third challenge, that of implementing analytics that have pedagogical and ethical integrity in a context where power and control over data are now of primary importance. We consider some of the concerns that learning analytics provoke, and suggest that Social Learning Analytics may provide ways forward. We conclude by revisiting the drivers and trends, and consider future scenarios that we may see unfold as SLA tools and services mature.

Keywords
Learning Analytics, Social Learning, Dispositions, Social Networks, Discourse, Informal Learning

Introduction
The concept of Learning Analytics is attracting significant attention within several communities with interests at the intersection of learning and information technology, including educational administrators, enterprise computing services, educators and learners. The core proposition is that, as unprecedented amounts of digital data about learners’ activities and interests become available, there is significant potential to make better use of this data to improve learning outcomes.

After introducing some of the conceptual roots of Learning Analytics (§2), we propose that the implementation of effective Social Learning Analytics is a distinctive part of this broader design space, and offers a grand challenge for technology-enhanced learning research and enterprise, in three important respects (§3).

1. The first is that the educational landscape is extraordinarily turbulent at present, in no small part due to technological drivers. The move to a participatory online culture sets a new context for thinking about analytics. Online social learning is emerging as a significant phenomenon for a variety of reasons, which we review (§4) in order to clarify the concept of online social learning (§5) and ways of conceiving social learning environments as distinct from other social platforms.

2. The second challenge is to understand the possibilities offered by different types of Social Learning Analytics, both those that are either inherently social (§6) and those that can be socialised, i.e., usefully applied in social settings (§7).

3. Thirdly, we face the challenge of implementing analytics that satisfy concerns about the limitations and abuses of analytics (§8).

We conclude (§9) by considering potential futures for Social Learning Analytics, if the drivers and trends reviewed continue.

Learning analytics
Learning analytics has its roots in two computing endeavours not specifically concerned with learning, but rather with strong business imperatives to understand internal organisational data, and external consumer behaviour.

• Business Intelligence focuses on computational tools to improve organisational decision-making through effective fusion of data collected via diverse systems. The earliest mention of the term ‘learning analytics’ that
we have found relates to business intelligence about e-learning products and services (Mitchell & Costello, 2000).

- **Data Mining**, also called Knowledge Discovery in Databases (KDD), is the field concerned with employing large amounts of data to support the discovery of novel and potentially useful information (Piatetsky-Shapiro, 1995). This field brings together many strands of research in computing, including artificial neural networks, Bayesian learning, decision tree construction, instance-based learning, logic programming, rule induction and statistical algorithms (Romero & Ventura, 2007).

From data mining developed the field of:
- **Educational Data Mining (EDM)** “an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in” (Baker & Yacef, 2009). Originally, relatively fine-grained, quantitative data came from private educational software applications—Romero and Ventura (2007) trace the first EDM publications to 1995—but their overview of the field shows that research projects multiplied after widespread adoption of virtual learning environments (VLEs) in the early 21st century.

Blackboard and Moodle are well-known examples of VLEs, which are also known as learning management systems (LMSs) and content management systems (CMSs). These tools automatically amass large amounts of log data relating to student activities. They not only record student activities and browse time, but also personal information such as user profiles, academic results, and interaction data. Many of them include student tracking capabilities as generic software features. Dawson (2009) reported that the depth of extraction and aggregation, reporting and visualisation functionality of these built-in analytics was often basic or non-existent, but in the last year, all of the major VLE products now include at least rudimentary analytics “dashboards.” Educational institutions have become increasingly interested in analysing the available datasets in order to support retention of students and to improve student results. This use of academic analytics stretches back for at least 50 years, but has become more significant in the last five years as datasets have grown larger and more easily available for analysis.

- **Academic Analytics** are described by Campbell & Oblinger (2007) as ‘an engine to make decisions or guide actions. That engine consists of five steps: capture, report, predict, act, and refine.’ They note that ‘administrative units, such as admissions and fund raising, remain the most common users of analytics in higher education today.’
- **Action Analytics** is a related term, proposed by Norris, Baer and Offerman (2009) to emphasise the need for benchmarking both within and across institutions, with particular emphasis on the development of practices that make them effective.

The Signals project at Purdue University is currently the field’s flagship example of the successful application of academic analytics, reporting significantly higher grades and retention rates than were observed in control groups (Arnold, 2010; Pistilli & Arnold, 2012). The project mines data from a VLE, and combines this with predictive modelling to provide a real-time red/amber/green traffic-light to students and educators, helping staff intervene in a timely manner where it will be most beneficial, and giving students a sense of their progress.

Encouraged by such examples, educational institutions are seeking both to embed academic/action analytics and to develop a culture that values the insights that analytics provide for organisational strategic planning and improved learner outcomes. A growing number of universities are implementing data warehouse infrastructures in readiness for a future in which they see analytics as a key strategic asset (Stiles, Jones, & Paradkar, 2011). These data warehouses store and integrate data from one or more systems, allowing complex queries and analysis to take place without disrupting or slowing production systems.

This brings us to the present situation; the first significant academic gathering of the learning analytics community was in 2011 at the 1st International Conference on Learning Analytics & Knowledge, doubling in size to 200 in 2012. The 2011 conference defined the term as follows:

> Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.
Clearly, this encapsulates strands from all the above fields, reflecting the topic’s interdisciplinary convergence but, in contrast to more theoretical research or artificial experimentation which might be published in some of the above fields, there is an emphasis on impacting authentic learning from real-world contexts, through the use of practical tools. There is also a shift away from an institutional perspective towards a focus on the concerns of learners and teachers. The main beneficiaries are no longer considered to be administrators, funders, marketing departments and education authorities, but instead are learners, teachers and faculty members (Long & Siemens, 2011).

The challenge of social learning analytics

In a literature analysis of the field, we found that in the discourse of academic analytics there is little mention of pedagogy, theory, learning or teaching (Ferguson, 2012). This reflects the roots of these analytics in management information systems and business intelligence, whose mission has been to guide strategic action by senior leaders in organisations, and whose tools deliver abstracted summaries of key performance indicators. In such contexts, senior executives do not have the time to delve into the process details of a particular individual’s or group’s interactions, and similarly, the arguments for academic analytics seem to focus on finding variables that predict positive or negative outcomes for cohorts of learners.

Performance indicators in educational settings typically involve outcomes-centric analytics based on learners’ performance on predefined tasks. Within formal education, success is typically defined as the display of expertise through summative assessment tasks (for example, assignments, exams or quizzes) intended to gauge mastery of discipline knowledge. The focus is on individual performance and on what has been achieved. This model is familiar within settings such as schools and universities, but it is less relevant in the context of online social learning, which involves lifelong learners drawing together resources and connections from across the Internet to solve real-life problems, often without access to the support of a skilled teacher or accredited learning institution.

Social Learning Analytics (SLA) are strongly grounded in learning theory and focus attention on elements of learning that are relevant when learning in a participatory online culture. They shift attention away from summative assessment of individuals’ past performance in order to render visible, and in some cases potentially actionable, behaviours and patterns in the learning environment that signify effective process. In particular, the focus of social learning analytics is on processes in which learners are not solitary, and are not necessarily doing work to be marked, but are engaged in social activity, either interacting directly with others (for example, messaging, friending or following), or using platforms in which their activity traces will be experienced by others (for example, publishing, searching, tagging or rating).

Social Learning Analytics is, we propose, a distinctive subset of learning analytics that draws on the substantial body of work demonstrating that new skills and ideas are not solely individual achievements, but are developed, carried forward, and passed on through interaction and collaboration. A socio-cultural strand of educational research demonstrates that language is one of the primary tools through which learners construct meaning. Its use is influenced by their aims, feelings and relationships, all of which shift according to context (Wells & Claxton, 2002). Another socio-cultural strand of research emphasises that learning cannot be understood by focusing solely on the cognition, development or behaviour of individual learners; neither can it be understood without reference to its situated nature (Gee, 1997; Wertsch, 1991). As groups engage in joint activities, their success is related to a combination of individual knowledge and skills, environment, use of tools, and ability to work together. Understanding learning in these settings requires us to pay attention to group processes of knowledge construction – how sets of people learn together using tools in different settings. The focus must be not only on learners, but also on their tools and contexts.

Viewing learning analytics from a social perspective highlights types of analytic that can be employed to make sense of learner activity in a social setting. This gives us a new way to conceive of both current and emerging approaches—as tools to identify social behaviours and as patterns that signify effective process in learning environments. Social Learning Analytics should render learning processes visible and actionable at different scales: from national and international networks to small groups and individual learners.

We turn now to review some of the features of the participatory online culture that drives this work.
The emergence of open, social learning

In this section, we identify some of the signals that many futures analysts and horizon-scanning reports on learning technology have highlighted as significant. Taken together, these create synergies that establish a radically new context for learning. In such a context, we argue, analytics focused on summative assessment of performance remain important but do not go far enough: we need to develop new sets of analytics that can be used to support learning and teaching in these new conditions.

We summarise these phenomena as:

- technological drivers
- the shift to ‘free’ and ‘open’
- demand for knowledge-age skills
- innovation requires social learning
- challenges to educational institutions.

Technological drivers

A key force shaping the emerging landscape is clearly the digital revolution. Only very recently do we have almost ubiquitous Internet access in wealthy countries and mobile access in many more. In addition, we now have user interfaces that have evolved through intensive use, digital familiarity from an early age, standards enabling interoperability and commerce across diverse platforms, and scalable computing architectures capable of servicing billions of real-time users and of mining the resulting data.

With the rise of social websites serving millions of users, such as Facebook, YouTube and Twitter, plus the thousands of smaller versions and niche applications for specific tasks and communities, we have witnessed a revolution in the ways in which people think about online interaction and publishing. Such social media platforms facilitate the publishing, indexing and tracking of user-generated media, provide simple-to-learn collaboration spaces, and enable social networking functions that are becoming ubiquitous: friending, following, messaging and status updates. Standards such as really simple syndication (RSS) allow information to be shared easily using structured data feeds, web services enable more sophisticated machine-machine interaction, and mobile devices expand the availability and localization of these services.

Internet services may also begin to apply pressure to one of the slowest evolving elements in educational provision: accreditation. Christensen et al. (2008) argue that the agencies controlling accreditation often stifle innovation and protect the status quo, because new approaches to learning/accreditation struggle to gain credibility unless they are associated with institutions that have the power to award established qualifications. However, as the infrastructure for secure identity management matures, and as the participatory, social culture fostered by Web 2.0 becomes more deeply ingrained in younger generations, initiatives such as OpenBadges may provide new ways to accredit learning outside established institutions. Moreover, as ubiquitous tools for capturing digital material make it easier to evidence learning and practical knowledge in authentic communities of practice, an e-portfolio of evidence might come to have equivalent or greater credibility than formal certificates.

However, changes in technology do not necessarily imply changes in pedagogy. Those who view education as information transfer will use interactive media for storage, drilling, testing and accessing information; those who seek conceptual change will seek to make use of their interactive qualities (Salomon, 2000). Technological shifts support analytics that draw on sets of big data—but they do not necessitate a shift towards analytics focused on such issues as conceptual change, distributed expertise, collaboration or innovation. So, if we do not accept simplistically that technology alone determines the future, we need to look elsewhere to understand the move towards online social learning and its associated analytics.

The shift to free and open

There has been a huge shift in expectations of access to digital content. The Internet makes possible completely new revenue-generation models due to the radically lower transaction costs incurred (compared to bricks and mortar
businesses with physical products) as one scales to hundreds of thousands of users. Andersen (2009) documents many ways in which online companies are able to offer quality services free of charge, producing an increasing expectation on the part of end-users of huge choice between free tools and sources of content hosted ‘in the cloud’.

Within education, the Open Education Resource (OER) movement has been a powerful vehicle for making institutions aware of the value of making high quality learning materials available, not only free of charge, but also in formats that promote remixing, in an effort to reap the benefits seen in the open-source software movement. This has not proven to be a simple matter, but OER has made huge progress, and is gaining visibility at the highest levels of educational policy.

This is amplified by efforts to make data open to machine processing as well as human interpretation. This requires not only a shift in mindset by data owners but also the construction of technological infrastructure to make it possible to publish data in useful formats. These efforts can be tracked within communities developing Linked Data and the Semantic Web, and their myriad applications communities, for example, Open Government, Open Mapping, Science 2.0 and Health 2.0.

Together, these very rapid shifts contribute to a new cultural context for the provision of learning services, in which the industrial-era value chain, previously delivered by a single institution, is disaggregated into smaller and smaller elements. The provision of content, community, tools and basic analytics may increasingly be expected to come free of charge, while learners may still consider paying for other services such as personalised learning journeys, personal tuition, career guidance, accreditation against formal standards and tailored analytics that support them on a variety of sites, not just within one institution.

**Demand for knowledge-age skills**

Technology is always appropriated to serve what people believe to be their needs and values. Since 1991, we have lived in the “knowledge age”—a period in which knowledge, rather than labour, land or capital, has been the key wealth-generating resource (Savage, 1996). This shift has occurred within a period when constant change in society has been the norm, and it is therefore increasingly difficult to tell which specific knowledge and skills will be required in the future (Lyotard, 1979). These changes have prompted an interest in “knowledge-age skills” that will allow learners to become both confident and competent designers of their own learning goals (Claxton, 2002).

Accounts of knowledge-age skills vary, but they can be broadly categorized as relating to learning, management, people, information, research/enquiry, citizenship, values/attributes and preparation for the world of work (Futurelab, 2007). From one viewpoint they are important because employers are looking for “problem-solvers, people who take responsibility and make decisions and are flexible, adaptable and willing to learn new skills” (Educational Subject Center, 2007, p. 5). More broadly, knowledge-age skills are related not just to an economic imperative but to a desire and a right to know, an extension of educational opportunities, and a “responsibility to realise a cosmopolitan understanding of universal rights and acting on that understanding to effect a greater sense of community” (Willinsky, 2005, p111). In both cases, there is a perceived need to move away from a curriculum based on a central canon of information towards learning that develops skills and competencies. This implies a need for ongoing analytics that can support the development of dispositions such as creativity and curiosity, collaboration skills and resilience.

**Innovation requires social learning**

The conditions for online social learning are also related to the pressing need for effective innovation strategy. In an accessible introduction to the literature and business trends, Hagel et al. (2010) argue that social learning is the only way in which organizations can cope in today’s fast-changing world. They invoke the concept of ‘pull’ as an umbrella term to signal some fundamental shifts in the ways in which we catalyse learning and innovation. They highlight quality of interpersonal relationships, tacit knowing, discourse and personal passion as key elements. This is a move away from having information pushed to us during spells of formal education towards a more flexible situation in which we pull resources and information to us as we need them. The move from “push” to “pull”
motivates analytics that can be accessed by learners at any point, employed in both informal and formal settings, are sensitive to social relationships, and build transferable learning dispositions and skills.

**Challenges to educational institutions**

Together, these forces create pressures on models of educational provision at all stages of education from childhood into workplace learning. Heppell (2007), amongst many, points to the need for an education system that helps people to help each other, rather than one that delivers learning. The barriers between formal and informal learning, and between online and face-to-face learning are currently being broken down, allowing the development of new models that take into account the range of learners’ experience outside formal study, and the affective elements of learning.

An example of this is Gee’s “affinity spaces,” which provide a model for online social learning and were first identified in video gaming environments. Affinity spaces are organized around a passion; within them, knowledge is both distributed and dispersed, they are not age graded, experts work alongside newcomers, learning is proactive but aided as people mentor and are themselves mentored, participants are encouraged to produce as well as to consume, smart tools are available to support learning and everyone, no matter what their level of experience or expertise, remains a learner (Gee, 2004, 2009). Other new models for learning are emerging from a variety of digital sources. Some examples amongst many are the learning affordances of the *World of Warcraft* online game, with its guilds and carefully planned, collectively executed strategies (Thomas & Brown, 2011), learners beginning to access and create knowledge through persistent avatar identities that can move between different environments (Ferguson, Sheehy, & Clough, 2010), and the development of distributed cognition within virtual worlds (Gillen, Ferguson, Peachey, & Twining, 2012).

These models suggest new ways of approaching learning analytics. Gee (2003) showed that well-designed video games incorporate analysis of the development of participants’ relevant knowledge and skills, so that their experience is constantly customized to their current level, effort and growing mastery, they are aware of ongoing achievements, and they are provided with information at the point when it can best be understood and used in practice.

Having noted some of the features of the emerging landscape for open, social learning, and the implications of these features for analytics, we now consider some of the key features of social learning, and the nature of online social learning environments.

**Characterising online social learning**

> Why has someone sawn down half of the beautiful cedar tree outside my office window? I can’t find this out from a book, and I don’t know anyone with the precise knowledge that I am looking for. It is as I engage in conversations with different people that my understanding of what I see outside my window increases, and I learn more about the tree’s history, health, ecosystem and future possibilities. It is not just the social construction of understanding that is important here, since this is a part of most human interactions. My intention to learn is part of what makes this social learning, as are interactions with others. This is not a one-sided engagement with books or online content—it involves social relationships. As such, it has lots of ‘affective’ aspects: people must be motivated to engage with me and I must have the confidence to ask questions in the first place, as well as some way of assessing the expertise of the people I’m talking to. (Ferguson, 2010)

Social learning has been conceptualised as societal learning in general, as processes of interaction that lead to concerted action for change, as group learning, and as the learning of individuals within a social context (Blackmore, 2010). Our conception of online social learning takes into account the changing affordances of a world in which social activity increasingly takes place at a distance and in mediated forms. It is succinctly expressed by Seely Brown and Adler (2008) as being “based on the premise that our understanding of content is socially constructed through conversations about that content and through grounded interactions, especially with others, around problems or actions.” Many others have, of course, argued for similar conceptions, unpacking this broad concept in great detail within the constructivist educational literature, and computer-supported collaborative learning (CSCL) research.
Social learning adds an important dimension to CSCL, introducing a particular interest in the non-academic contexts in which it may take place (including the home, social network, and workplace) and the use of free, ready-to-hand online tools, with no neatly packaged curriculum or signed-up peer cohort, no formally prescribed way to test one’s understanding and no pre-scheduled activities (Blackmore’s (2010) edited readings remind us how far back everyday, non-digital social learning goes in learning theory, and provide us with foundations for extension into the digital realm).

While OERs greatly increase the amount of good quality material available online to learners, another consequence can be that individual learners find themselves adrift in an ocean of information, struggling to solve ill-structured problems, with little clear idea of how to solve them, or how to recognise when they have solved them. At the same time, distributed networks of learners are grappling with ‘wicked problems’ such as climate change, which offer the same challenges on a grander scale. Social learning infrastructure could have a key role to play in these situations, helping learners connect with others who can provide emotional and conceptual support for locating and engaging with resources, just as in our tree story at the start of this section. This forces us to ask whether our current educational and training regimes are fit for purpose in equipping our children, students and workforce with the dispositions and skills needed under conditions of growing uncertainty—a challenge explored in detail by many others, for example in the collection edited by Deakin Crick (2009).

The Open University, where we are based, has been seeking to address these issues with its SocialLearn project, aimed at supporting large-scale social learning. In the early days of the project, Weller (2008) identified six broad principles of SocialLearn: Openness, Flexibility, Disruptive, Perpetual beta, Democracy and Pedagogy. Following a series of workshops, Conole (2008) proposed a set of learning principles for the project—thinking & reflection, conversation & interaction, experience & interactivity and evidence & demonstration—and articulated how these could be linked to characteristics of social learning.

Distilling this array of perspectives, we have derived a simple working definition focused on three dynamics, which serves to guide us in designing for meaningful interpersonal and conceptual connection:

Online social learning can take place when people are able to:

• clarify their intention—learning rather than browsing
• ground their learning—by defining their question/problem, and experimenting
• engage in learning conversations—increasing their understanding.

A significant feature of the Web 2.0 paradigm is the degree of personalisation that end-users now expect. However, a me-centred universe has self-evident limitations as a paradigm for holistic development: learning often disorients and reorients one’s personal universe. User-centred is not the same as Learner-centred: what I want is not necessarily what I need, because my grasp of the material, and of myself as a learner, is incomplete. The centrality of good relationships becomes clear when we remind ourselves that a university’s job is to teach people to think, and that deeper learning requires leaving a place of cognitive and emotional safety where assumptions are merely reinforced—see the extensive research on learning dispositions that characterize this readiness (for example, Claxton, 2001; Perkins, Jay, & Tishman, 1993). This implies challenge to stretch learners out of their comfort zones, underlining the importance of affirmation and encouragement that give a learner the security to step out.

As Figure 1 shows, the design of a social media space tuned for learning involves many alterations and additions to a generic space for social media. Within an online space tuned for learning, friends can become learning peers and mentors, informal endorsements are developed into verifiable accreditation, information exchanges become learning conversations and, likewise, generic web analytics need to be developed into learning analytics that can be used in such an environment.

To summarise: we have outlined what we mean by online social learning, some of the major drivers that help to explain why it is emerging as a phenomenon, and some of the elements that may differentiate a social learning environment from other social media spaces. We have also indicated why these factors require new approaches to learning analytics. Constructivist pedagogies suggest the need for a shift away from a positivist approach to analytics and towards analytics that are concerned with conceptual change, distributed expertise, collaboration and innovation. This ties in with an increasing emphasis on knowledge-age skills and their associations with learning dispositions.
such as creativity and resilience. Within an open environment, there is a need for a range of analytics that can extend beyond an institutional platform in order to provide support for lifelong learners at all points in their learning journey. These learners may be organised in classes and cohorts, but they may also need analytics that help them to learn together in looser groupings such as communities and networks. These analytics, and their associated recommendations, will be informed by those developed for social media tools and platforms, but they will be tuned for learning, examples being prompting the development of conversations into educational dialogue, recommending resources that challenge learners to leave their comfort zones, or making learners aware that social presence and role are increasingly important to attend to in a complex world.

![Figure 1. Dimensions of the social learning design space](image)

Together, these motivate a conception of Social Learning Analytics as a distinctive class of analytic.

**Inherently social learning analytics**

Social learning analytics make use of data generated by learners’ online activity in order to identify behaviours and patterns within the learning environment that signify effective process. The intention is to make these visible to learners, to learning groups and to teachers, together with recommendations that spark and support learning. In order to do this, these analytics make use of data generated when learners are socially engaged. This engagement includes both direct interaction—particularly dialogue—and indirect interaction, when learners leave behind ratings, recommendations or other activity traces that can influence the actions of others. Another important source of data consists of users’ responses to these analytics and their associated visualizations and recommendations.

We identify two inherently social analytics, and three socialised analytics:

- **Social Network Analytics**—interpersonal relationships define social platforms and link learners to contacts, resources and ideas.
- **Discourse Analytics**—language is a primary tool for knowledge negotiation and construction.

Socialised analytics—although these are relevant as personal analytics, they have important new attributes in a collective context:

- **Content Analytics**—user-generated content is one of the defining characteristics of Web 2.0
• **Disposition Analytics**—intrinsic motivation to learn lies at the heart of engaged learning and innovation

• **Context Analytics**—mobile computing is transforming access to people, content and both formal and informal learning.

We do not present these as an exhaustive “taxonomy,” since this would normally be driven by, for instance, a specific pedagogical theory or technological framework in order to motivate the category distinctions. We are not grounding our work in a single theory of social learning, nor do we think that a techno-centric taxonomy is helpful. These categories of analytics respond to the spectrum of drivers reviewed above, drawing on diverse pedagogical and technological underpinnings as reviewed above, and further cited below as we introduce each category. We summarise the essence of each approach, identify examples of tools, and then consider how these tools are being, or might be, used to support online social learning. In this section, we introduce the two inherently social analytics.

**Social network analytics**

*Essence of social network analysis*

Networked learning involves the use of ICT to promote connections between one learner and other learners, between learners and tutors, and between learning communities and learning resources (Jones & Steeples, 2003). These networks are made up of actors (both people and resources) and the relations between them. Actors with a relationship between them are said to be tied and these ties can be classified as strong or weak, depending on their frequency, quality or importance (Granovetter, 1973). Social network analysis is a perspective that has been developed to investigate the network processes and properties of ties, relations, roles and network formations, and to understand how people develop and maintain these relations to support learning (Haythornthwaite & de Laat, 2010).

Fortunato (2010) describes social networks as “paradigmatic examples of graphs with communities”; social network analysis brings graph theory from the field of mathematics together with work on interpersonal and communal relationships from the fields of sociology and communication. The many uses of social network analysis applicable to social learning include detection of communities within networks (Clauset, Newman, & Moore, 2004; Fortunato, 2010); identification of types of subset within a network where a level of cohesion exists and depends on properties such as proximity, frequency and affinity or other properties (Reffay & Chanier, 2003); investigation of the density of social networks (Borgatti, Mehra, Brass, & Labianca, 2009); and exploration of individuals’ centrality within a network (Wasserman & Faust, 1994).

*Social network analysis tools*

Many tools have been developed to support social network analysis in the context of learning. Commercial products such as Mzinga can be used to identify learners with the highest and most active participation in a network, those who are having the most influence on the activity of others and those who have the potential to make most impact. SNAPP (Social Networks Adapting Pedagogical Practice) is a freely available network visualisation tool that re-interprets discussion forum postings as a network diagram. These diagrams can be used to trace the growth of course communities, to identify disconnected students, to highlight the role of information brokers and to visualise how teacher support is employed within the network (Bakharia & Dawson, 2011; Dawson, Bakharia, & Heathcote, 2010).

Gephi is a free, open-source platform that supports visualisation and exploration of all kinds of networks. In an extended series of blog posts, Hirst has explored ways in which this tool can be used to explore the learning networks that develop around shared resources and online course. His work picks out different networks with interconnected interests, identifies the interests that are shared by actors in a network, and highlights not only the role played by information brokers in sharing resources, but also the roles played by resources in connecting networks.

*Network-focused social learning analytics*

Social network analysis is a useful tool for examining online learning because of its focus on the development of interpersonal relationships, and its view that technology forms part of this process. It thus offers the potential to
identify interventions that are likely to increase the potential of a network to support the learning of its actors by linking them to contacts, resources and ideas.

Haythornthwaite and De Laat (2010) approach this form of analysis from two perspectives: egocentric and whole network. Egocentric networks are described from the point of view of the individual, who is set at the centre of an array of relationships both formally and informally connected with learning. Studying networks in this way can help to identify the people from whom an individual learns, where conflicts in understanding may originate, and which contextual factors influence learning. A whole-network view, on the other hand, considers the distribution of information and the development of learning across a set of people. In this case, analysis can characterise the network in terms of its character, interests and practices. This whole-network view is able to take “the results of pairwise connections to describe what holds the network together” (Haythornthwaite & de Laat, 2010, p. 189).

Characterising the ties between actors adds a different dimension to this analysis—people rely on weak ties with people they trust when accessing new knowledge or engaging in informal learning, but make use of strong ties with trusted individuals as they deepen and embed their knowledge (Levin & Cross, 2004). Another option is to combine social network analysis with content analysis and context analysis to gain a richer picture of networked learning, investigating not only who is talking to whom, but what they are talking about and why they are talking in this way (De Laat, Lally, Lipponen, & Simons, 2006; Hirst, 2011).

As social network analysis is developed and refined, it has the potential to be combined with other social learning analytics in order to define what counts as a learning tie and thus to identify which interactions promote the learning process. It also has the potential to be extended in order to take more account of interactions with resources, identifying indirect relationships between people which are characterised by their interaction with the same resources rather than through direct communication.

Social learning discourse analytics

Essence of discourse analysis

Discourse analysis is the collective term for a wide variety of approaches to the analysis of series of communicative events. Some of these approaches cannot easily be employed as online social learning discourse analytics because they focus on face-to-face or spoken interactions and may require intensive examination of semiotic events from a qualitative perspective. Others provide new ways of understanding the large amounts of text generated in online courses and conferences. Schrire (2004) used discourse analysis to understand the relationship between the interactive, cognitive and discourse dimensions of online interaction, examining initiation, response and follow-up (IRF) exchanges. More recently, Lapadat (2007) has applied discourse analysis to asynchronous discussions between students and tutors, showing how groups of learners create and maintain community and coherence through the use of discursive devices.

Discourse analysis tools

There are many tools available for the online analysis of text and discourse; the Digital Research Tools Wiki currently lists 55. These range from well-known visualisation tools such as Wordle and Tag Crowd to powerful generic tools such as NVivo, which can be used to support a range of qualitative research methods.

A method of discourse analysis that relies heavily on electronic tools and computer processing power is corpus linguistics, the study of language based on examples of real-life use. The corpus of examples is typically in electronic form and may be massive; the European Corpus Initiative Multilingual Corpus includes 98 million words covering most of the major European languages, while the British National Corpus is a 100-million-word sample of a wide range of written and spoken sources. Automated software, such as WMatrix, facilitates quantitative investigation of such corpora (O’Halloran, 2011).
A different approach to seeking to extract structure from naturally occurring but relatively unstructured texts is to ask users to add more structure themselves. This is an extension of asking users to enrich resources with metadata, which we see in social tagging. Learners cannot be asked to structure their annotations on documents and contributions to discussion simply to facilitate computational processing, since there would be no value for them in doing so. However, significant research in concept mapping (Novak, 1998) and computer-supported argumentation (Scheuer, Loll, Pinkwart, & McLaren, 2010) has shown that this can be a pedagogically effective discipline to ask of students in a formal academic context, and within organisational contexts, the mapping of conversations can promote quality meetings and shared ownership of outcomes amongst diverse stakeholders (Selvin & Buckingham Shum, 2002).

Cohere is a web-based tool that provides a medium not only for engaging in structured online discourse, but also for summarizing or analysing it (Buckingham Shum, 2008). Following the approach of structured deliberation/argument mapping, Cohere renders annotations on the web, or a discussion, as a network of rhetorical moves: users must reflect on, and make explicit, the nature of their contribution to a discussion. This tool can be used to augment online conversation by making explicit information on the rhetorical function and relationship between posts. Users also have the option to browse their online dialogue as a semantic network of posts rather than as a linear text.

**Discourse-focused social learning analytics**

A sociocultural perspective on learning “highlights the possibility that educational success and failure may be explained by the quality of educational dialogue, rather than simply in terms of the capability of individual students or the skill of their teachers” (Mercer, 2004, p. 139). The ways in which learners engage in dialogue are indicators of how they engage with other learners’ ideas, how they compare those ideas with their personal understanding, and how they account for their own point of view, which is an explicit sign of the stance they hold in the conversation. Mercer and his colleagues distinguished three social modes of thinking that are used by groups of learners in face-to-face settings: disputational, cumulative and exploratory talk (Mercer, 2000; Mercer & Littleton, 2007). Disputational dialogue is characterised by disagreement and individualised decision-making; in cumulative dialogue speakers build on each other’s contributions but do not critique or challenge these. Exploratory dialogue is typically regarded as the most desirable by educators because speakers share knowledge, challenge ideas, evaluate evidence and consider options together.

Learning analytics researchers have built on this work to provide insight into textual discourse in online learning (Ferguson, 2009), providing a bridge to the world of online learning analytics for knowledge building. Initial investigations (Ferguson & Buckingham Shum, 2011) suggest that indicators of exploratory dialogue—challenges, extensions, evaluations and reasoning—can be automatically identified within online discussion. This analysis can be used to provide recommendations about relevant learning discussions, as well as to prompt the development of meaningful learning dialogue.

The Cohere structured deliberation platform has been extended by De Liddo and her colleagues (2011) to provide learning analytics that identify:

- **Learners’ attention**—what they focus on, which problems and questions they raise, which comments they make and which viewpoints they express
- **Learners’ rhetorical attitude to discourse contributions**—areas of agreement and disagreement, the ideas supported by learners and the ideas questioned by learners
- **Distribution of learning topics**—the people who propose and discuss the most contentious topics
- **Learners’ relationships**—beyond the undifferentiated ties of social network analysis, Cohere users are tied with semantic relationships (such as supporting or challenging), showing how learners relate to each other and how they act within a discussion group.

While informal text chat is difficult to analyse automatically in any detail, due to non-standard use of spelling, punctuation and grammar, more formally structured texts such as a journal article can be analysed using natural language processing technologies. Sándor and her colleagues (Sándor, Kaplan, & Rondeau, 2006; Sándor & Vomdran, 2009) have used the Xerox Incremental Parser (XIP) to highlight key sentences in academic articles in order to focus an evaluator’s attention on the key rhetorical moves within the text which signal claims to contribute
to knowledge. Analysis of XIP and human annotation suggests that they are complementary in nature (Sándor, De Liddo, & Buckingham Shum, 2012).

Whitelock and Watt analysed discourse using Open Mentor, a tool for teachers to analyse, visualise and compare the quality of their feedback to students (Whitelock & Watt, 2007, 2008). Open Mentor uses a classification system based on that of Bales (1950) in order to investigate the socio-emotive aspects of dialogue as well as the domain level. A standard charting component is then used to provide interactive bar chart views onto tutors’ comments, showing the difference between actual and ideal distributions of different comment types. Tutors can use these analytics to reflect on their feedback, and the analytics can also be used to recommend moves towards the types of feedback that students find most useful.

The development of the field of learning analytics has brought approaches to discourse that originated in the social sciences more closely in contact with statistical methods of extracting and representing the contextual usage and meaning of words (Landauer, Foltz, & Laham, 1998). A social learning analytics perspective offers the possibility of harnessing these methods and understandings in order to provide analytics and representations that can help learners to develop their conversations into reasoned arguments and educational dialogue.

Socialised learning analytics

Discourse and social network analytics are inherently concerned with social interaction. In the context of learning, they already have a strong focus on the learning group. In this section, we consider three kinds of learning analytic that are more typically viewed from the perspective of the isolated learner who may be making no use of interpersonal connections or social media platforms. We argue that these analytics take on significant new dimensions in the context of online social learning.

Social learning disposition analytics

Essence of learning dispositions

The first of these socialised learning analytics is the only one of our five categories that originated in the field of educational research rather than being adapted to apply to the analysis of learning. A well-established research programme has identified, theoretically, empirically and statistically, a seven-dimensional model of learning dispositions (Deakin Crick, 2007). These dispositions can be used to render visible the complex mixture of experience, motivation and intelligences that make up an individual’s capacity for lifelong learning and influence responses to learning opportunities (Deakin Crick, Broadfoot, & Claxton, 2004). They can be used to assess and characterise the complex mixture of experience, motivation and intelligences that a learning opportunity evokes for a specific learner. It is these developing qualities that make up an individual’s capacity for lifelong learning (Deakin Crick, et al., 2004).

Learning dispositions are not “learning styles,” a blanket phrase used to refer to a wide variety of frameworks that have been critiqued on a variety of grounds, including lack of contextual awareness (Coffield, Moseley, Hall, & Ecclestone, 2004). By contrast, important characteristics of learning dispositions are that they vary according to context, and that focused interventions have been shown to produce statistically significant improvements in diverse learner groups, ranging in age from primary school to adults, demographically from violent young offenders and disaffected teenagers to high achieving pupils and professionals, and culturally from middle-class Western society to Indigenous communities in Australia (Buckingham Shum & Deakin Crick, 2012).

Together, learning dispositions comprise the seven dimensions of “learning power”: changing & learning, critical curiosity, meaning making, dependence & fragility, creativity, relationships/interdependence and strategic awareness (Deakin Crick, 2007). Dynamic assessment of learning power can be used to reflect back to learners what they say about themselves in relation to these dimensions, and to provide teachers with information about individuals and groups that can be used to develop students’ self-awareness as well as their ownership of and responsibility for their learning.
Disposition analysis tools

The ELLI (Effective Lifelong Learning Inventory) assessment tool arose from an exploratory factor analytic study involving 2000 learners. Since then, it has been developed in a range of educational settings worldwide as an instrument to help assess capacity for lifelong learning (Deakin Crick, 2007; Deakin Crick et al., 2004; Small & Deakin Crick, 2008). ELLI is a self-report questionnaire which individuals are asked to answer with a specific piece of recent learning in mind. These responses are used to produce a learning profile, a graphical representation of how the learner has reported themselves in relation to the dimensions of learning power: “very much like me,” “quite like me” or “a little like me.” This diagram is not regarded as a description of fixed attributes but as the basis for a mentored discussion with the potential to spark and encourage changes in the learner’s activities, attitude and approach to learning.

In order to gather ELLI data globally, with quality and access controls in place, and to generate analytics fast enough to impact practice in a timely manner, ELLI is hosted within a learning analytics infrastructure called the Learning Warehouse. This supports large-scale analysis of international datasets (e.g., >40,000 ELLI profiles), providing portals to organisations including remote Australian communities, schools in China, Malaysia, Germany, Italy, US, and corporates in the UK (Buckingham Shum & Deakin Crick, 2012).

Disposition-focused social learning analytics

Learning dispositions are personal, related to the identity, personhood and desire of the learner (Deakin Crick & Yu, 2008). They can be regarded as socialised learning analytics when the emphasis shifts away from the learner as individual towards the learner in a social setting. From this perspective, two elements of disposition analytics are particularly important—their central role in an extended mentoring relationship, and the importance of relationships / interdependence as one of the seven key learning dispositions.

The ELLIment tool provides a collaboration space for a learner and mentor to reflect on a learner’s ELLI profile, and agree on interventions. EnquiryBlogger mines information from a blogging tool set up to support enquiry, providing learners and teachers with visual analytics reflecting student activity and their self-assessment of progress in their enquiry, use of learning dispositions, and overall enjoyment. This then enables appropriate and timely intervention from teachers and, being a blogging environment, comments from peers (Ferguson, Buckingham Shum, & Deakin Crick, 2011).

Mentors play an important part in social learning, providing both motivation and opportunities to build knowledge. They may act as role models, encouraging and counselling learners, and can also provide opportunities to rehearse arguments and to increase understanding (Anderson & Shannon, 1995; Ferguson, 2005; Liu, Macintyre, & Ferguson, 2012). People providing these online support relationships may be able to provide more useful assistance if they are aware of the prior knowledge, progress and goals of the person asking a question (Babin, Tricot, & Mariné, 2009).

From a social learning perspective, disposition analytics provide ways of stimulating conceptual change, distributed expertise, collaboration and innovation. They tie in with an increasing emphasis on knowledge-age skills, and can be used to encourage learners to reflect on their ways of perceiving, processing and reacting to learning interactions. From the perspective of teachers and mentors, awareness of these elements contributes significantly to their ability to engage groups of learners in meaningful, engaging education.

Social learning content analytics

Essence of content analytics

Whereas disposition analytics have been developed within the field of education, content analytics have only recently been associated with education, originating in technical fields concerned with recommender systems and information retrieval (Drachsler, Hummel, & Koper, 2008; Zaïane, 2002). Content analytics is used here as a broad heading for the variety of automated methods that can be used to examine, index and filter online media assets, with
the intention of guiding learners through the ocean of potential resources available to them. Note that these analytics are not identical to content analysis, which is concerned with description of the latent and/or manifest elements of communication (Potter & Levine-Donnerstein, 1999). Combined with learning context analytics or with defined search terms, content analytics may be used to provide recommendations of resources that are tailored either to the needs of an individual or to the needs of a group of learners.

Research in information retrieval represents the leading edge of techniques for the automated indexing and filtering of content, whether textual, or multimedia (for example, images, video, or music). The state of the art in textual and video information retrieval tools is displayed annually in the competitions hosted at the Text Retrieval Conference (see Little, Llorente, & Rüger, 2010 for a review). Visual similarity search is an example of multimedia content analysis that uses features of images such as colour, texture and shape in order to find material that is visually related. This allows near-duplicate detection, known object identification and general search. Together, these elements can be used to provide novel methods of suggesting, browsing or finding educational media.

Other approaches to content analytics are more closely aligned with content analysis. These involve examination of the latent elements that can be identified within transcripts of exchanges between people learning together online. This method has been used to investigate a variety of issues related to online social learning, including collaborative learning, presence and online cooperation (de Wever, Schellens, Vallcke, & van Keer, 2006). These latent elements of interpersonal exchanges can also be used to support sentiment analysis, using the objectivity/subjectivity of messages, and the emotions expressed within them to explore which resources are valued, and the motivations behind recommendations (Fakhraie, 2011).

Content analysis tools

Web-based search engines are the default tools to which most learners and educators turn for text search, but multimedia search is becoming increasingly possible. While some approaches exploit the metadata around a multimedia asset, such as the text surrounding a photo, rather than analyse its actual content, true image-based search on the web is now available (for instance, Google Image search allows the filtering of results by colour). Some e-commerce websites enable product filtering by visual similarity, and mobile phone applications are able to parse images such as book covers, in order to retrieve their metadata (e.g., http://www.snaptell.com).

Turning to transcript analysis, commonly used tools for content analysis include NVivo and Atlas.ti, both of which are software packages designed to support the analysis of unstructured information and qualitative data. However, these are manual tools for human analysts. Erkens and Janssen (2008) review the challenges of automated analysis, and describe Multiple Episode Protocol Analysis (MEPA), which has been validated against human coders, and used to automatically annotate chat transcripts from learning environment in numerous studies. In the selection of any of these tools, researchers face the bigger challenge of identifying an analytic framework that “emphasizes the criteria of reliability and validity and the counting of instances within a predefined set of mutually exclusive and jointly exhaustive categories” (de Wever et al., 2006). The validity of content analysis of online discussion has been persistently criticised (Pidgeon, 1996, p. 78) and it has proved difficult to identify empirically validated content analysis instruments to use in these contexts (Rourke, Anderson, Garrison, & Archer, 2003).

Content-focused social learning analytics

How do these tools take on a new dimension in social learning? Visual Similarity Search can be used to support navigation of educational materials in a variety of ways, including discovering the source of an image, finding items that share visual features and may provide new ways of understanding a concept, or finding other articles, talks or movies in which a given image or movie frame is used (Little, Ferguson, & Rüger, 2011). Content analytics take on a social learning aspect when they draw upon the tags, ratings and additional data supplied by learners. An example is iSpot, which helps learners to identify anything in the natural world (Clow & Makriyannis, 2011). When a user first uploads a photo to the site, it has little to connect it with other information. The addition of a possible identification by another user ties that photo to other sets of data held externally in the Encyclopaedia of Life and within the UK’s National Biodiversity Network. In the case of iSpot, this analysis is not solely based on the by-products of
interaction, an individual’s reputation within the network helps to weight the data that is added. The site’s reputation system has been developed with the purpose of magnifying the impact of known experts. Overall, the example of iSpot suggests one way in which content analytics can be combined with social network analytics to support learning. The two forms of analytics can also be used to support the effective distribution of key resources through a learning network.

Another approach is to apply content analysis to the interplay of learning activities, learning objects, learning outcomes, and learners themselves, establishing semantic relations between different learning artefacts. This is the approach taken by LOCO-Analyst, which is used to analyse these semantic relations and thus provide feedback for content authors and teachers that can help them to improve their online courses (Jovanović et al., 2008). This type of analysis can draw on the information about user activity and behaviour that is provided by tools such as Google Analytics and userfly.com as well as by the tools built into environments such as Moodle and Blackboard.

Social learning context analytics

**Essence of context analytics**

Overall, social learning analytics can be applied to a wide variety of contexts that extends far beyond institutional systems. They can be used in formal settings such as schools, colleges and universities, in informal contexts in which learners choose both the process and the goal of their learning (Vavoula, 2004) and by mobile learners in a variety of situations (Sharples, Taylor, & Vavoula, 2005). In some cases, learners are in synchronous environments, structured on the basis that participants are co-present in time, and at others they are in asynchronous environments, where the assumption is that they will be participating at different times (Ferguson, 2009). They may be learning alone, in a network, in an affinity group, in communities of inquiry, communities of interest or communities of practice (Ferguson, 2009). Here we are grouping under the heading “context analytics” the various analytic tools that expose, make use of or seek to understand these contexts in relation to learning.

Zimmerman and his colleagues (2007) provide a definition of context that allows the definition of the context of an entity (for example, a learner) depending on five distinct categories:

- **Individuality context** includes information about the entity within the context. In the case of learners, this might include their language, their behaviour, their preferences and their goals.
- **Time context** includes points in times, ranges and histories so can take into account work flow, long-term courses and interaction histories.
- **Location context** can include absolute location, location in relation to people or resources, or virtual location (IP address).
- **Activity context** is concerned with goals, tasks and actions.
- **Relations context** captures the relations of an entity with other entities, of example with learners, teachers and resources.

Early work in context-aware computing treated the environment as a shell encasing the user and focused on scalar properties such as current time and location, together with a list of available objects and services (see, for example, Abowd, Atkeson, Hong, Long, & Pinkerton, 1997; Want, Hopper, Falcao, & Gibbons, 1992). The focus was on the individual user receiving data from an environment rather than interacting with it. This model did not acknowledge the dynamics of interaction between people and the environment. When considered in the context of learning, it did not provide information that could help people to modify their environment in order to create supportive workspaces or form social networks with those around them or accessible online (Brown et al., 2010).

**Context analysis tools**

The MOBIlearn project took a different view, considering context to be a dynamic process, constructed through learners’ interactions with learning materials and the surrounding world over time (Beale & Lonsdale, 2004; Syvänen, Beale, Sharples, Ahonen, & Lonsdale, 2005). The MOBIlearn context awareness subsystem was developed to allow learners to maintain their attention on the world around them while their device presents content, options
and resources that support their learning activities. The developers of the system designed the system to analyse a variety of available data in order to produce learner-focused information and recommendations, taking into account not only the physical environment but also social relationships.

Environmental information such as geographical position allows us to provide location-specific information, e.g., for a museum. Other user information such as the identification and presence of another person allows us to create a peer-to-peer network for informal chat. But the combination of the two may allow us to determine that the other user is a curator, and we can provide the mechanisms for one to give a guided tour to the other. (Beale & Lonsdale, 2004)

The Active Campus tool was another one developed to prompt connections with learners and resources. The aim was to provide a tool that could analyse people, resources and events in the vicinity and then act like a pair of “x-ray glasses,” providing opportunities for serendipitous learning by letting users see through crowds and buildings to reveal nearby friends, potential colleagues and interesting events (Griswold et al., 2004).

**Context-focused social learning analytics**

The MOBIlearn project produced several recommendations to be considered in the design process of an adaptive and pervasive learning environment. Some of these are focused on the physical design of tools, but others are directly relevant to the development of context-focused social learning analytics, specifically:

- Organizing the information provided to the user according to the availability for cooperation (students), advice (experts, instructors) and groups available at a given moment.
- Supporting the communication between users by presenting tools, such as news groups and chats, ordered by their current popularity in the learning community (placing first the most popular, or the most relevant to the learner according to the profile, at any given moment).
- Encouraging users to cooperate and affiliate by pushing the information when relevant opportunities occur. Actions by the system are guided, for example, by the information related to a group-based modeling that takes into account each user’s evident interest in certain piece(s) of information (Syvänen et al., 2005).

These suggest fruitful ways forward in this area. In the case of online learning, context analytics can draw upon readily available data such as profile information, timestamps, operating system and location. Such data mining can support recommendations that are appropriate for learners’ situation, the time they have available, the devices they can access, their current role and their future goals. Context analytics can also be used to highlight the activity of other learners in a community or network, through tag clouds, hash tags, data visualizations, activity streams and emergent folksonomies.

In addition to development work in this field, there is also a need for substantial theoretical work that can underpin it. Social network analysts have spent many years identifying elements and structures that have been found to support learning and which can be used to create contexts that promote the development of sophisticated learning networks. There are currently no such sophisticated analytics available to help us develop suitable contexts for other groupings known to support social learning, such as affinity groups and communities of practice. We also lack the long-term analytics of learner behaviour that could help us to analyse context in order to support the development of personal learning narratives, learning trajectories or other understandings of lifelong learning (Gee, 2004; Jones & Preece, 2006; Lipman, 2003; Wenger, 1998).

**The challenge of powerful analytics**

Having explained how we are conceiving social learning analytics, we now consider some of the critiques around the balance of power in learning analytics, in response to which we will conclude by sketching potential future scenarios that may address these concerns.

New forms of measurement and classification—for that is essentially what learning analytics are—are rightly exposed to disciplinary and ethical critiques concerning issues such as: who is defining the measures, to what ends,
what is being measured, and who has access to what data? In their incisive critique of classification systems, Bowker and Star (2000) demonstrate how these become the mechanisms by which we choose not only how to remember, but also systematically forget, what is known. If a phenomenon is not visible within a classification scheme, it is systematically erased. The issue of power is, therefore, a central one to confront.

This dilemma sits at the heart of the controversy around any policy dependent on a predefined performance indicator. Schools, universities, faculties or individuals whose work is invisible within a classification scheme are disenfranchised when defined by powerful stakeholders with associated rewards/sanctions. Whether this is reasonable sparks debate as to whether phenomena are being justifiably ignored because they are not something to be encouraged, or whether it is simply that they are too hard to quantify for automated processing and performance grading.

The challenge for learning analytics is more complex still. As described above, at least some forms of learning analytics research have an interest in using data generated by users as a *by-product* of online activity (for example, asking/answering questions, or recommending resources), rather than as an intentional form of evidence of learning (such as taking a test or submitting an essay). Building on this potentially noisy data, research into recommendation engines goes one step further, exploring the potential to mine such data for patterns that can be acted on by software agents in some way—perhaps in the form of feedback to learners via a personal analytics dashboard or as modifications to the content that is displayed based on the system’s model of the learner. Such research must engage fully with questions around the academic, pedagogical and ethical integrity of the principles for defining such patterns and recommender algorithms, and who is permitted to see them within the set of stakeholders.

Important concerns (boyd & Crawford, 2011) are beginning to be expressed about learning analytics, such as the following variants on longstanding debates at the intersection of education, technology and artificial intelligence:

- Analytics are dependent on computational platforms that use, re-use and merge learner data, both public and private: institutions should steer clear of open data and minimise the merging of datasets of any sort until there are much clearer ethical and legal guidelines.
- Analytics could disempower learners, making them increasingly reliant on institutions providing them with continuous feedback, rather than developing meta-cognitive and learning-to-learn skills and dispositions.
- Analytics are a crude way to operationalise proxy measures of teacher effectiveness, and will be used to compare and contrast student outcomes, leading to the gaming of the system: “learning and teaching to the analytic” to maintain performance indicators that do not genuinely promote meaningful learning.

In sum, learning analytics and recommendation engines are always designed with a particular conception of “success,” thus defining the patterns deemed to be evidence of progress, and hence, the data that should be captured. A marker of the health of the learning analytics field will be the quality of debate around what the technology renders visible and leaves invisible.

Briefly, let us consider how these issues may be seen through a Social Learning Analytics lens, recognising that a more detailed treatment is needed in future work. If the values and practices we see in the open, social web inform the ways in which SLAs are deployed, we may see ways to address these concerns. For example:

- If SLA tools and data are placed in the hands of learners, the balance of power shifts significantly. When the exposure of personal data to analytics is voluntary, when a group’s data is collectively owned, and when gaming the system or trying to pretend to be someone you are not incurs social sanctions, the risks of abuse are arguably lower than when a hierarchical institution carries the unrealistic burden of responsibility for controlling a living ecosystem of participants, data and tools. It is realistic to note that the above imply a maturing in technologies, learner literacies, and institutional practices around the management of personal data, compared to the situation we have today.
- If analytics are drawing learners’ attention to their development as self-aware, intrinsically motivated learners, they are being moved in the opposite direction to becoming passively dependent on the institution or platform to tell them how they are doing and what to do next.
- If analytics are focused on providing formative feedback to improve learning process, rather than making automated judgments about mastery levels in a given subject, there might be fewer concerns around the removal of human mentors from the feedback loop. We also hypothesise that the risks of “gaming the analytic” reduce: SLA activity patterns are by definition hard to fabricate privately, so not only are learners fooling themselves if
they fake behaviour (e.g., designed to look like skillful discourse, supportive networking, or self-reflection), they risk making fools of themselves among peers for whom authenticity and trustworthiness are valued personal qualities.

Conclusion: SLA future scenarios

Let us conclude by engaging in the early stages of what Miller (2007) terms “futures literacy”—stretching our imaginations in disciplined ways in order to sketch potential futures, were social learning analytics to develop in line with these cultural shifts. Consider the forces identified earlier (§4), and for each, imagine future scenarios in which SLA values, tools and practices have matured beyond today’s nascent state.

The digital infrastructure is reaching a state of maturity that enables non-technical people to engage with expertly designed “walk up and use” interfaces on both large-screen and mobile devices, to connect with people and information on a global scale, and to make their contributions via social media platforms.

- **Potential SLA future:** Institutions lacking the infrastructure needed for computationally intensive analytics and recommendation engines will call on SLA services in the computing “cloud,” following the business developments we are now seeing to offer commercial learning analytics cloud services on school/university data. Individual learners or communities who need such services also utilise these services. Some companies and educational institutions will exploit their pedagogical expertise to provide SLA consulting services. As we see the commercialization of the analytics computing space, there is an argument that at this point the field needs a complementary Open Learning Analytics innovation platform (SoLAR, 2011).

“Free and Open” is a key expectation and dynamic within online social learning. It highlights the recalibration that is taking place around expectations of freely provided quality services, accompanied by readiness to pay for value-added services once the free service has proven itself. Data is expected to be accessible, appropriately licensed for remixing and, wherever possible, in machine-readable formats to facilitate interoperability and avoid data or users being locked into a given platform.

- **Potential SLA future:** Many SLA tools become available in open source versions, making them customisable within the myriad unique social contexts in which they may be deployed. It becomes normal that SLA patterns and data are open, shareable resources for reflection, and analysis in alternative tools. In addition to a diverse palette of free SLA tools, an economy grows which helps learners to configure these to create meaningful toolkits that support particular kinds of learning, or work well with particular platforms. Learners are willing to pay for more powerful features, once the most successful tools have earned their right to charge. A key lesson from the social web paradigm, and a long-held aspiration of researchers into end-user customisability, is that when empowered with appropriately flexible tools, an ecosystem grows in which new roles are created for different kinds of user to customise their tools (MacLean, Carter, Lovstrand, & Moran, 1990).

Aspirations across cultures have been shifting in empirically verifiable ways towards a growing desire for participation and self-expression. The social web is an expression of this shift, providing a significant medium for many people to construct their identity.

- **Potential SLA future:** The outputs of SLA tools become an important part of individuals’ sense of identity, and their ability to evidence their skills. For example, we might see Badges such as: “I am a good broker between communities,” “I can distill complex debates into their essence,” “I can mentor learners in building their creativity.”

Innovation in complex, turbulent environments requires social knowledge-creation and negotiation infrastructures built on quality relationships and conversations—beyond impersonal “transactions”—in order for individuals, groups and organisations to be agile enough to respond to turbulent change and to work together to solve “wicked problems.”

- **Potential implications for SLA:** SLAs become an integral part of the employee’s toolkit, helping to track the swirl of people, conversations and resources by rendering significant changes in coherent ways that keep cognitive load at a manageable level, rather than amplifying demands on attention.
The role of educational institutions is changing. They are moving increasingly to provide personalised support for learning how to think deeply, and learning how to be an effective member of the communities that one cares about.

- Potential implications for SLA: Educational institutions are no longer the only option for evidencing advanced learning. Analytics become a new form of trusted evidence, being generated from verifiable public datasets, or private datasets that could not have been reasonably fabricated, such as a reputable online community.

In sum, if it is the case that these tectonic shifts define a new context for thinking about learning, in particular around questions of power and the central role of interpersonal relationships, by extension they set a new context for thinking about learning analytics. They call into question the assumption inherited from the business intelligence and management information systems orientation, that learning analytics are designed and controlled primarily by institutional educators and administrators in order to optimize learners’ performance, and hence the institution’s performance. This is not at all to argue that academic/action analytics are unimportant—but it now becomes clear that this is only one of a range of possible analytics scenarios.

To conclude, we have motivated the concept of Social Learning Analytics as a response to some of the forces reshaping the educational landscape, and our growing understanding that many forms of learning most relevant to becoming a citizen in our complex society are socially grounded and evidenced phenomena. SLAs may be deployed as institutional tools in conventional courses, to yield insight for educators and administrators. Equally, however, they should be seen as tools to be placed in the hands of the very subjects being analysed—the learners—and for the many informal learning contexts that we now see outside the walls of conventional institutions. It would indeed be ironic if the ways in which Social Learning Analytics tools were deployed did not honour and promote the open, democratising, critical dynamics that underpin much of the participatory, social web philosophy—dynamics which SLA tools make visible in new ways.

Acknowledgements
We gratefully acknowledge The Open University for resourcing the SocialLearn Project, several anonymous reviewers for their constructive reviews on earlier drafts, and the encouragement from researchers and practitioners who have found these ideas valuable in their own work.

References


Integrating Data Mining in Program Evaluation of K-12 Online Education

Jui-Long Hung*, Yu-Chang Hsu and Kerry Rice
Department of Educational Technology, Boise State University, Boise, USA // andyhung@boisestate.edu // hsu@boisestate.edu // krice@boisestate.edu

*Corresponding author

ABSTRACT
This study investigated an innovative approach of program evaluation through analyses of student learning logs, demographic data, and end-of-course evaluation surveys in an online K–12 supplemental program. The results support the development of a program evaluation model for decision making on teaching and learning at the K–12 level. A case study was conducted with a total of 7,539 students (whose activities resulted in 23,854,527 learning logs in 883 courses). Clustering analysis was applied to reveal students’ shared characteristics, and decision tree analysis was applied to predict student performance and satisfaction levels toward course and instructor. This study demonstrated how data mining can be incorporated into program evaluation in order to generate in-depth information for decision making. In addition, it explored potential EDM applications at the K-12 level that have already been broadly adopted in higher education institutions.

Keywords
Educational data mining, Program evaluation, K–12 virtual school, Pattern discovery, Predictive modeling

Introduction
Traditionally, the majority of online instructors and institutional administrators rely on web-based course evaluation surveys to evaluate online courses (Hoffman, 2003). The data and information are then used to help inform online program effectiveness and generate information for program-level decision-making. While it enjoys wide use, the survey method only provides learners’ self-report data, not their actual learning behaviors.

Several studies have found self-reported data were not consistent with actual learning behaviors (Hung & Crooks, 2009; Picciano, 2002). This inconsistency can potentially compound the already problematic lack of direct observation opportunities. Online program administrators need more effective tools to provide customized learning experiences, to track students’ online learning activities for overseeing courses (Delavari, Phon-annualusuk, & Beikzadeh, 2008), to depict students’ general learning characteristics (Wu & Leung, 2002), to identify struggling students (Ueno, 2006), to study trends across courses and/or years (Hung & Crooks, 2009), and to implement institutional strategies (Becker, Ghedini, & Terra, 2000). Each of these needs can be addressed by mining educational data. Nowadays, various educational data are stored in database systems. This is especially true for online programs, wherein student learning behaviors are recorded and stored in Learning Management Systems (LMS). Program administrators can take advantage of emerging knowledge and skills by extracting and interpreting those data. The purpose of this study is to propose a program evaluation framework using Educational data mining.

Program evaluation
Program evaluation is the means by which a program assures itself, its administration, accrediting organizations, and students that it is achieving the goals delineated in its mission statement (Nichols & Nichols, 2000). Evaluation can be done by a variety of means. The most common form of evaluation is through surveying students regarding courses/faculty/programs (e.g., Cheng, 2001; Hoffman, 2003; Spirduso & Reeve, 2011). However, making causal inferences based on a one-time assessment is risky (Astin & Lee, 2003). Nevertheless, perceptional survey data cannot accurately reflect real learning behaviors (Hung & Crooks, 2009; Picciano, 2002). Although various scholars (e.g., Grammatikopoulos, 2012; Vogt & Slish, 2011) have proposed systematic frameworks (e.g., interviews and observation) in order to obtain objective knowledge via multiple means, these methods are difficult to implement in a fully online program.
Educational data mining

Data mining (DM) is a series of data analysis techniques applied to extract hidden knowledge from server log data (Roiger & Geatz, 2003) by performing two major tasks: Pattern discovery and predictive modeling (Panov, Soldatova, & Dzeroski, 2009). Educational data mining (EDM) is a field which adopts data mining algorithms to solve educational issues (Romero & Ventura, 2010). Romero & Ventura (2010) reviewed 306 EDM articles from 1993 to 2009 and proposed desired EDM objectives based on the roles of users. For the purpose of this study, which is designed to inform administrators, the list is limited to objectives for administrators:

- Enhance the decision processes in higher learning institutions
- Streamline efficiency in the decision making process
- Achieve specific objectives
- Suggest certain courses that might be valuable for each class of learners
- Find the most course effective way of improving retention and grades
- Select the most qualified applicants for graduation
- Help to admit students who will do well in higher education settings

Based on the theory of bounded rationality, decision-making is a fully rational process of finding an optimal choice given the information available (Elster, 1983). An ideal program evaluation framework should provide multiple facets of information to decision makers. Therefore, integrating more than one data source and analytic method is essential for an effective program evaluation.

Figure 1. Program evaluation framework
Program evaluation framework

Figure 1 shows the framework of the proposed program evaluation method. The core strategy of this framework is data triangulation (Jick, 1979) which combines multiple data sources (learning logs, course evaluation survey, and demographic data) and multiple methods (pattern discovery and predictive modeling) to generate accurate, in-depth results. Using this framework, the authors conducted a program evaluation case study to evaluate how the proposed program evaluation framework can support administrators’ decision making.

Method

Data source

In this case study, data were collected from a statewide K–12 online institution that serves over 16,000 students in a northwestern state in the U.S. The institution provides fully online courses to K–12 students. Courses were designed by subject-matter curriculum designers and subject-matter teachers to standardize course materials. Teachers were required to complete an online orientation prior to teaching courses for the institution. Teachers received the same or similar training for online teaching provided by the institution. Site coordinators are located at each district in the state and regional principals oversee teacher evaluation. The following data were collected for the academic year of 2009-2010 (3,604 students enrolled in Fall 2009 and 3,935 students in Spring 2010): 1) LMS activity logs; 2) student demographic data; and 3) course evaluation survey data. All data tables were stored in the database and interconnected with unique identifiers (e.g., course ID).

LMS activity logs

The LMS activity logs were collected from the Blackboard activity accumulator (Blackboard Inc., 2010) for the Fall 2009 and Spring 2010 academic terms. The following records were removed in data preprocessing: irrelevant fields (e.g., group ID), irrelevant records (e.g., login failure), and data stored in wrong or mismatched fields (about 11.8% of overall activity logs). After data preprocessing, a total of 23,854,527 activity logs were collected from 7,539 students in 883 courses. These students took 1 to 18 courses in the 2009–2010 academic year.

Student demographic data

The following demographic data were collected for data analysis, including age, gender, graduation year, city, school district, number of online course(s) taken, number of online course(s) passed, number of online course(s) failed, and final grade average.

Course evaluation survey

A course evaluation survey investigated students’ satisfaction toward their course and instructor. Course satisfaction contained eight questions related to course content, five related to course structure, and eleven related to instructor satisfaction. Records containing any missing values were removed from the analysis. In addition, because student identifiers were not collected during the Fall 2009 survey implementation, which prevents the researchers from associate survey responses with demographic data and LMS activity logs, only Spring 2010 survey data (2618 respondents) were analyzed in this study.

Engagement level

Engagement is considered to be a key variable for enabling and encouraging learners to interact with the material, with the instructor, and with one another, as well as for learning in general. In this study, engagement level was measured by the frequency of various learning interactions that happened within the LMS. Variables under the category “Student Engagement Variable” in Table 1 were applied to measure each student’s engagement level, which included:
- Average frequency of logins per course.
- Average frequency of tab accessed per course (if the course was organized using “tabbed” navigation).
- Average frequency of module accessed per course (if the course was organized using “modules”).
- Average frequency of clicks per course.
- Average frequency of course accessed per course (from Blackboard portal to course site).
- Average frequency of page accessed per course (content created using the page tool). The Page tool allows instructors to include files, images, and text as links on the course menu
- Average frequency of course content accessed per course (content created using the content tool). The Content tool allows instructors to create course content within the content area.
- Average number of discussion board entries per course.

Variables

Table 1 lists variables collected from Blackboard, the student demographic database, and the course evaluation survey. Some variables were transformed with calculations in order to generate more meaningful variables for analysis. For example, student’s birth year was transformed to age. The summary of all learning activities was aggregated to a new variable called “frequency of clicks” that represents each student’s total frequency of clicks in the Blackboard LMS. If students took more than one course during the analysis period, variables of learning activities (e.g., frequency of total clicks and frequency of course access), performance (e.g., final grade), and survey (e.g., course satisfaction and instructor satisfaction) were averaged.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Descriptions</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>stuID</td>
<td>Student’s ID</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>Age</td>
<td>Student’s age</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>City</td>
<td>Student’s residential city</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>District</td>
<td>Student’s school district</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>Grade_Avg</td>
<td>Average course grade</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Click_Avg</td>
<td>Average frequency of clicks/course</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Content_Access_Avg</td>
<td>Average frequency of course content</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Course_Access_Avg</td>
<td>Average frequency of course accessed/course</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Page_Access_Avg</td>
<td>Average frequency of page accessed/course</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>DB_Entry_Avg</td>
<td>Average number of discussion board</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Tab_Access_Avg</td>
<td>Average frequency of tab accessed/course</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Login_Avg</td>
<td>Average frequency of logins/course</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Module_Avg</td>
<td>Average frequency of module accessed/course</td>
<td>Student Engagement Variable</td>
</tr>
<tr>
<td>Gender</td>
<td>Gender</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>HSGradYear</td>
<td>High school graduation year</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>School</td>
<td>Student’s school</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>No_Course</td>
<td>Number of course taken</td>
<td>Student Demographic Variable</td>
</tr>
<tr>
<td>No_Fail</td>
<td>Number of course failed</td>
<td>Student Performance Variable</td>
</tr>
<tr>
<td>No_Pass</td>
<td>Number of course passed</td>
<td>Student Performance Variable</td>
</tr>
<tr>
<td>Pass rate</td>
<td>Average individual student pass rate for all courses in academic year 2009-2010 (&gt;= 0 and &lt;=1)</td>
<td>Student Performance Variable</td>
</tr>
<tr>
<td>cSatisfaction_Avg</td>
<td>Average course satisfaction including 8 questions related to course content and 5 questions related to course structure</td>
<td>Student Perception Variable</td>
</tr>
<tr>
<td>iSatisfaction_Avg</td>
<td>Average instructor satisfaction including 11 questions related to instructor.</td>
<td>Student Perception Variable</td>
</tr>
</tbody>
</table>
Analytic tools

SAS Enterprise Miner 6.1 (SAS Institute Inc., USA) was employed to perform the following data mining tasks in this study: 1) student clustering which describes shared characteristics of students who passed or failed their course; 2) perception and performance predictions which identify key predictors of course satisfaction, instruction satisfaction, and final grade. Because one of the major target audiences of this article is K–12 administrators, the authors utilized methods like decision tree and K-means clustering, which can produce results more intuitive for non-data miners.

Results

Student clustering

Clustering analysis

K-means algorithm (Hartigan & Wongm, 1979; Budayan, Dikmen, & Birgonul, 2009) was applied to group students based on their shared characteristics (Internal Standardization = Range; Maximum Number of Clusters = 6). Total clusters were limited to avoid trivially small or exclusive groups, the identification of which was outside the purposes of this case study. A pass rate equal to “1” means a student passed all courses during the period of analysis. A pass rate equal to “0” means a student failed all courses during the period of analysis. A pass rate between “0” and “1” means a student passed some, but not all, courses during the period of analysis. In clustering analysis, pass rate was set up as the standard for classification and six clusters were generated.

Table 2 includes the results of clustering analysis in academic year 2009-2010. The following are shared characteristics of each cluster.

- **Cluster 1** (316 students, pass rate = 55.07%, all males): Cluster 1 consists of students who are older than Cluster 3 to 6. They were lower-engaged than Cluster 5 and 6 but higher than Cluster 3 and 4. On average, each student took 2.76 courses and failed about half of them.
- **Cluster 2** (320 students, pass rate = 56.11%, all females): Similar to Cluster 1, Cluster 2 consists of students who are older than Clusters 3 to 6. They are lower-engaged than Cluster 5 and 6 but higher than Cluster 3 and 4. On average, each student took 3.03 courses and failed about half of the courses.
- **Cluster 3** (594 students, pass rate = 0%, all males): Cluster 3 and 4 includes the lowest-engaged students. Cluster 3 students are all male. On average, each student took 1.43 courses and failed all of them.
- **Cluster 4** (601 students, pass rate = 0%, all females): Cluster 4 includes the lowest-engaged female students. On average, each student took 1.39 courses and failed all of them.
- **Cluster 5** (2,311 students, pass rate = 100%, all males): Cluster 5 and 6 represent the highest-engaged students. Cluster 5 students are all male. On average, each student took 1.59 courses and passed all of them.
- **Cluster 6** (3,397 students, pass rate = 100%, all females): Cluster 6 represents the highest-engaged female students. On average, each student took 1.64 courses and passed all of them.

<table>
<thead>
<tr>
<th>Variables</th>
<th>CL1</th>
<th>CL2</th>
<th>CL3</th>
<th>CL4</th>
<th>CL5</th>
<th>CL6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Students</td>
<td>316</td>
<td>320</td>
<td>594</td>
<td>601</td>
<td>2311</td>
<td>3397</td>
</tr>
<tr>
<td>Pass rate = 0</td>
<td>0</td>
<td>0</td>
<td>594</td>
<td>601</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pass rate (&gt; 0 and &lt; 1)</td>
<td>316</td>
<td>320</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pass rate = 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2311</td>
<td>3397</td>
</tr>
<tr>
<td>GenderF</td>
<td>316</td>
<td>0</td>
<td>594</td>
<td>0</td>
<td>2311</td>
<td>0</td>
</tr>
<tr>
<td>GenderM</td>
<td>16.91</td>
<td>17.06</td>
<td>16.69</td>
<td>16.82</td>
<td>16.6</td>
<td>16.59</td>
</tr>
<tr>
<td>Grade_Avg</td>
<td>50.11</td>
<td>52.82</td>
<td>22.44</td>
<td>20.85</td>
<td>81.75</td>
<td>85.4</td>
</tr>
<tr>
<td>Click_Avg</td>
<td>583.15</td>
<td>549.09</td>
<td>440.17</td>
<td>416.4</td>
<td>892.49</td>
<td>881.69</td>
</tr>
<tr>
<td>Content_Access_Avg</td>
<td>112.96</td>
<td>112.2</td>
<td>93.78</td>
<td>89.43</td>
<td>180.34</td>
<td>177.96</td>
</tr>
<tr>
<td>Course_Access_Avg</td>
<td>170.26</td>
<td>172.29</td>
<td>133.94</td>
<td>141.52</td>
<td>281.5</td>
<td>284.22</td>
</tr>
<tr>
<td>DB_Entry_Avg</td>
<td>4.08</td>
<td>5.28</td>
<td>2.78</td>
<td>4.22</td>
<td>8.28</td>
<td>9.57</td>
</tr>
<tr>
<td>Login_Avg</td>
<td>29.4</td>
<td>24.35</td>
<td>23.58</td>
<td>19.18</td>
<td>47.92</td>
<td>46.42</td>
</tr>
<tr>
<td>Module_Access_Avg</td>
<td>156.18</td>
<td>145.02</td>
<td>112.7</td>
<td>102.38</td>
<td>249.16</td>
<td>240.79</td>
</tr>
</tbody>
</table>
The clusters generated from cluster analysis were associated with two geographical variables: city and school district, in order to examine whether certain types of students were from specific areas. Differences in engagement were found depending on location. Clusters 1 to 6 had similar geographical distributions except for three larger cities (populations larger than 100,000). Cluster 5 (all male, pass rate = 100%) included a larger group of students from one large city. Cluster 6 (all female, pass rate = 100%) included a larger group of students from the other two large cities. There is no notable difference of school district distributions across clusters.

Findings

Findings below were summarized from the clustering analysis.
1) Students with higher engagement levels usually had higher performance.
2) Younger students (CLs 5 & 6) who lived in larger cities were more successful than those in smaller cities (CLs 3 & 4) and older students (CLs 1 & 2).
3) All-failed students who were also low-engaged consisted of approximately 15.9% on average per course.
4) All-passed students who were also high-engaged consisted of approximately 75.7% students on average per course.
5) Based on Cluster 1 and 2, on average, older students (age > 16.91) tended to take more than two courses with pass rates ranging from 54.09-56.11%.
6) On average, high-engaged students demonstrated engagement levels twice that of low-engaged students.
7) Frequencies of reading behaviors (such as content access and page access) were much higher than discussion behaviors (p<0.001).
8) Female students were more active than male students in online discussions (with higher DB_Entry avg frequency).
9) Female students had higher pass rates than male students.

Average clicks per course in different subject areas

Table 3 shows students’ average frequencies of total clicks and performances per course in different subject areas. Total clicks were equal to the summarized frequency of overall learning activities. The results show that Math and Science had the highest number of total clicks per course and of total clicks per student per course. However, for those who took Math and/or Science courses, their average final grades (56.70 and 64.41 accordingly) were lower than the overall final grade average (71.11). This indicates students participated actively in courses of these two subject areas, but they failed to achieve expected outcomes (70 or higher). Possible reasons for this outcome could be related to course design and/or best practice in teaching strategies for Math and Science courses. On the other hand, English courses received a lower number of clicks combined with less than expected outcomes. Encouraging motivation and engagement in these courses could have a profound effect on future outcomes. Students in Foreign Language and Health not only participated in learning activities actively, but also obtained the highest grades, on average, in each of these two subject areas.

<table>
<thead>
<tr>
<th>Subject Areas</th>
<th>Total Clicks</th>
<th>Total Clicks/student</th>
<th>No of Students</th>
<th>Final Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drivers Ed</td>
<td>4,808.00</td>
<td>227.97</td>
<td>21.09</td>
<td>78.40</td>
</tr>
<tr>
<td>Electives</td>
<td>5,353.63</td>
<td>247.69</td>
<td>21.61</td>
<td>76.59</td>
</tr>
<tr>
<td>English</td>
<td>4,807.79</td>
<td>239.98</td>
<td>20.03</td>
<td>62.09</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>7,824.63</td>
<td>439.40</td>
<td>17.81</td>
<td>76.54</td>
</tr>
<tr>
<td>Health</td>
<td>6,641.80</td>
<td>269.99</td>
<td>24.60</td>
<td>83.58</td>
</tr>
<tr>
<td>Math</td>
<td>7,898.35</td>
<td>444.05</td>
<td>17.79</td>
<td>56.70</td>
</tr>
<tr>
<td>Science</td>
<td>9,015.16</td>
<td>603.53</td>
<td>14.94</td>
<td>64.41</td>
</tr>
</tbody>
</table>
Findings

10) Subjects where the level of activity was effective and consistent with student outcomes included Driver Education, Electives, Foreign Language, Health, and Social Studies.

11) Subjects where the level of activity was inconsistent with student outcomes included Math, Science and English. Math and Science courses had high activity levels with less than expected outcomes.

12) Subjects where the level of activity was low and consistent with low student outcomes included English.

Subject preferences

Figure 2 shows percentages of female and male students in different subject areas. Because the original female versus male ratio is 1.34, subject preferences for female and male students were revealed by comparing female/male ratio in each subject with the original ratio. Subjects above the dashed line are those with higher female ratios.

Findings

13) Female students preferred taking Electives, Foreign Language, and Social Studies.


Pass rate in different subject areas

Table 4 consists of two parts. The first part examines whether pass rates of female and male students in different subjects have significant differences. “F vs. M” compares gender pass rate difference using t-tests. The second part examines pass rate difference between Fall 2009 and Spring 2010 within the same gender. For example, “F vs. F” compares pass rate difference between Fall 2009 and Spring 2010 female students in difference subjects by using t-tests. Numbers marked with asterisks represent differences that have statistical significance.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Fall 2009</th>
<th>Spring 2010</th>
<th>Fall 2009 vs. Spring 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drivers Ed</td>
<td>91%</td>
<td>89%</td>
<td>.623</td>
</tr>
<tr>
<td>Electives</td>
<td>85%</td>
<td>79%</td>
<td>.000*</td>
</tr>
</tbody>
</table>

*Numbers marked with asterisks represent differences that have statistical significance.*
<table>
<thead>
<tr>
<th>Subject</th>
<th>F (Spring)</th>
<th>M (Fall)</th>
<th>p-value</th>
<th>F (Spring)</th>
<th>M (Fall)</th>
<th>p-value</th>
<th>F (Spring)</th>
<th>M (Fall)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>69%</td>
<td>57%</td>
<td>.000*</td>
<td>83%</td>
<td>72%</td>
<td>.000*</td>
<td>83%</td>
<td>72%</td>
<td>.000*</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>83%</td>
<td>80%</td>
<td>.088</td>
<td>91%</td>
<td>89%</td>
<td>.121</td>
<td>91%</td>
<td>89%</td>
<td>.121</td>
</tr>
<tr>
<td>Health</td>
<td>92%</td>
<td>90%</td>
<td>.023*</td>
<td>97%</td>
<td>97%</td>
<td>.058</td>
<td>97%</td>
<td>97%</td>
<td>.058</td>
</tr>
<tr>
<td>Math</td>
<td>55%</td>
<td>54%</td>
<td>.769</td>
<td>72%</td>
<td>70%</td>
<td>.761</td>
<td>72%</td>
<td>70%</td>
<td>.761</td>
</tr>
<tr>
<td>Science</td>
<td>71%</td>
<td>67%</td>
<td>.223</td>
<td>85%</td>
<td>81%</td>
<td>.418</td>
<td>85%</td>
<td>81%</td>
<td>.418</td>
</tr>
<tr>
<td>Social Studies</td>
<td>79%</td>
<td>74%</td>
<td>.000*</td>
<td>89%</td>
<td>84%</td>
<td>.000*</td>
<td>89%</td>
<td>84%</td>
<td>.000*</td>
</tr>
</tbody>
</table>

**Note:** Statistical significance refers to the possible accurate rate of a statement by testing it with statistical methods. “*p < .05” means the statement is at least 95% accurate (error rate is less than 5%). **p < .001**

**Findings**

15) Overall, females significantly performed better than male students, especially in the following subject areas: Electives, English, and Social Studies.

16) The fail rates during the Fall 2009 term was significantly higher than those during the Spring 2010 term, especially in those subjects with higher fail rates such as English, Math, Science, and Social Studies. After these results were revealed the researchers subsequently learned from the administrators of the program reported in this case study that they had adopted an early alarm system in Spring 2010 to track all communications between instructors and students. The results show those strategies could have improved students’ pass rates in most subject areas.

**Student performance and engagement by course number**

Due to the previous results indicating that students in Math, Science, and English had lower performance than those in other subject areas, researcher were interested in identifying potential anomalies within this group which might help to explain the reasons for the results. Further analysis was applied to identify which Math, Science, and English courses resulted in the highest performance and which Math, Science, and English courses resulted in the lowest performance.

Researchers divided courses into three conditions: (a) high-engaged, high-performance, (b) high-engaged, low performance, and (c) low-engaged, low-performance based on student behaviors within the course. Courses categorized as high-engaged and high-performance might represent courses with both effective design and effective implementation because students were highly engaged and achieved expected outcomes. Those categorized as high-engaged and low-performance might represent courses with less effective course design because students were unable to achieve expected outcomes despite what appears to be effective implementation. Finally, courses categorized as low-engaged and low performance might represent courses with less effective course design and less effective course implementation. Our analysis revealed that regardless of the content area, most high-engaged, low performance, or low-engaged, low performance courses were entry-level courses. Most high-engaged, high performance courses were advanced level courses.

Students’ responses to the survey question asking students to indicate their reasons for taking an online course were then incorporated to help further interpret the results. The majority of responses from young students enrolled in courses categorized as high-engaged and high-performing were “the course was not available in my school.” The majority of older student responses in courses that were categorized as low-engaged and low performance were “I was making up a class I had failed.”

**Findings**

17) Regardless of Math, Science, or English subject-matter, entry level courses tended to have lower performance whether students were categorized as low-engaged or high-engaged. This may speak more to course structure, design, and support than to the effectiveness of instruction.

18) The reasons students enrolled in a course may influence their engagement level and performance. Student survey responses indicated that students who retook courses they have previously failed, tended to demonstrate lower engagement and lower performance. If students took courses which were not available in their schools, these students were usually high-engaged and high performing.
Predictive analysis

CRT Decision Tree analysis (Breiman, Friedman, Olshen, & Stone, 1984) was applied to construct predictive models combining course related data and survey results (Splitting Criterion: Gini; Leaf Size: 60; Maximum Depth: 10; Assessment Measure: Average Squared Error). These settings allow for a larger sequence of sub-trees in order to enrich the study’s findings. Decision Trees classifies instances by sorting them down the tree from the root to the leaf nodes. In the tree structures, leaf nodes represent classifications, and branches represent conjunctions of features that lead to different target values. The following three variables were adopted as dependent variables in the Decision Tree analysis: 1) Average course grade; 2) average course satisfaction; and 3) average instructor satisfaction. Survey questions on course satisfaction and instructor satisfaction can be retrieved from http://goo.gl/x8j18

- Average course grade is each student’s final course grade (range: 0-100). If a student took more than one course, average course grade is the average of multiple courses.
- Average course satisfaction was generated by averaging the scores from eight survey questions related to course content and the five survey questions related to course structure (range: 1–7). If a student took more than one course, average course satisfaction is the average of satisfaction scores from multiple courses.
- Average instructor satisfaction was generated by averaging the scores from 11 survey questions related to the instructor satisfaction (range: 1–7). If a student took more than one course, average instructor satisfaction is the average of satisfaction scores from multiple courses.

Final grade prediction

All variables in Table 1 were imported for final grade prediction. Average course grade was used as the dependent variable and the remainders were treated as independent variables. Because the tree results contained too much information, blank nodes were used to represent the results excluded from the data interpretation. Figure 3 shows the decision tree for final grade prediction. In academic year 2009-2010, 75.7% of students passed all courses, 15.9% of students failed all courses, and 8.4% passed some but not all of their courses. The left branch of the decision tree represents students who passed all courses. The results indicate a positive correlation between engagement level and performance (higher engaged => higher performance). The right branch of the decision tree represents students who had failed in one or more courses. The results imply a negative correlation between engagement level and performance (lower engaged => lower performance).

![Figure 3. Final grade prediction (complete chart: http://goo.gl/NIfWu)](http://goo.gl/NIfWu)

Findings

19) Engagement level and gender have stronger effects on student final grades than age, school district, school, and city. For most students, high engaged => high performance.
20) Compared with other Blackboard components such as discussion board entries and content access, tab access has negative effects on student performance (higher tab accessed => lower performance).

21) Female students performed better than male students.

**Final grade prediction (external variables)**

Additional decision tree analysis was conducted to investigate how external variables (i.e., non-learning activity variables) influenced student performance. Figure 4 is a portion of the decision tree for academic year 2009-2010.

![Figure 4](http://goo.gl/B8AvB)

**Findings**

22) Based on the predictive model, female students performed better than male students.

23) Students who were around 16 years old or younger performed better than those who were 18 years or older.

![Figure 5](http://goo.gl/5NLWl)

**Course satisfaction prediction** (complete chart: http://goo.gl/5NLWl)
Satisfaction prediction

Decision tree analysis was also conducted to predict students’ satisfaction levels toward their course and instructor. Fall 2009 survey data could not be associated with variables in Blackboard, so the following results are limited to Spring 2010 only.

Course satisfaction

All the scores calculated from the responses to survey questions on course satisfaction were averaged into the scores of one course satisfaction variable. The value of “7” for this variable represents highest satisfaction with a course and “1” represents lowest satisfaction with a course. Figure 5 is a portion of the decision tree regarding course satisfaction.

Findings
24) Students with higher average final grades (> 73.25, with a maximum score of 100) had higher course satisfaction.
25) Students who passed all courses or passed some of their courses had higher course satisfaction than all-failed students.
26) Students who took two or more courses in Spring 2010, whether they passed those courses or not, had higher course satisfaction.
27) Female students had higher course satisfaction than male students.
28) Online behaviors (i.e., frequency of page accessed and number of discussion board entries) had minor effects on course satisfaction (higher frequency/number => higher course satisfaction).
29) Students in different cities showed different course satisfaction levels.

Instructor satisfaction

All the scores calculated from the responses to survey questions on instructor satisfaction were averaged into the scores of one instructor satisfaction variable. The value of “7” for this variable represents highest satisfaction with an instructor and “1” represents lowest satisfaction with an instructor. Figure 6 is a portion of the decision tree regarding instructor satisfaction.

Findings
30) Students with higher average final grades (> 73.25, with a maximum score of 100) indicated higher instructor satisfaction.
31) Students who took two or more courses in Spring 2010, whether they passed those courses or not, showed higher instructor satisfaction.

32) Female students indicated higher instructor satisfaction than male students.

33) Online behaviors (frequency of module accessed) had minor effects on instructor satisfaction (higher frequency => higher course satisfaction). However, there were six students indicated low instructor satisfaction, despite extremely high frequency of course access and high final grades.

34) Older students taking one course (> 17.5 years old) had higher instructor satisfaction.

35) Students from different schools showed different satisfaction levels for their online instructors.

36) Younger female students (<15.5 years old) with lower average final grade (<76.5) indicated lower instructor satisfaction.

Discussion

This study is a first attempt at program evaluation combining multiple data sources. The goal of this project was to propose a new program evaluation framework in order to generate sufficient information for program-level decision-making. The advantages of this framework are data triangulation and data interpretation. Below are triangulation and interpretation results:

- Female students generally performed better than male students (findings (9), (15), (21), and (22)); however, the findings were limited to the following subject areas: Electives, English, and Social Science (finding (15)).
- High-engaged students generally performed better than low-engaged students (findings (1), (10), (12) and (19)); however, the findings were limited to non-STEM courses (findings (10) and (11)). One possible factor influencing high-engaged students’ inability to consistently achieve expected outcomes may have been poor course design.
- Younger students generally performed better than older students (findings (2) and (23)); however, the findings were limited to students in larger cities (findings (2)).
- One possible explanation for older students’ generally lower performances may be that older students took more than two courses per semester for credit recovery (findings (5) and (18)). Younger students took fewer courses, but the fact the courses were generally not available in school district may have increased motivation (findings (2) and (18)).

Overall, using multiple forms of data allows for a more meaningful analysis of actual student behaviors, and the identification of potential relationships with demographic data, satisfaction data, and student outcomes. The result is a much richer and deeper analysis of student performance and teaching, as well as of effective course design, than could ever be accomplished with survey data or behavior mining alone.

Demographic and performance

Based on results revealed by the program evaluation framework, some indicators can be applied to identify students more likely to be successful and those more likely to be at-risk. In this study, a student who possessed more of the characteristics listed below was more likely to be successful:

- Female
- Younger than 16.5 years
- Took one or two courses per semester
- Took a Foreign Language or Health course
- Lived in a larger city

A student who possessed more of the characteristics below was more likely to be at risk of failure:

- Male
- Older than 18 years
- Took more than two courses per semester
- Took entry-level courses in Math, Science, or English
- Lived in a smaller city
These indicators can be applied to develop an early warning system (Macfadyen & Dawson, 2010), so administrators and teachers can have a list of successful and at-risk students before each semester starts.

**Engagement and performance**

Based on data mining analysis, higher-engaged students usually had higher performance. This finding is also supported by previous studies (e.g., Hung & Crooks, 2008; Hung & Zhang, 2009). However, the conclusion may be limited to courses which were well designed and implemented. In this study, entry-level courses tended to have lower performance, regardless as to whether students were categorized as low-engaged or high-engaged. This means high-engaged students might still have lower final grades if they were in a course with course structure, design, and/or support issues.

Lim & Morris (2009), when studying post-secondary students, found junior and senior students had significantly higher survey mean scores in perceived learning, learning application, and learning involvement than freshman and sophomore students. Assuming higher perceived learning, learning application, and learning involvement equates to high motivation, the authors could not explain why older students had significantly higher engagement than young students. Our study, by combining analysis of engagement and performance through data mining with survey responses, revealed why students had different levels of engagement and performance. For example, the majority of responses from the students enrolled in courses categorized as high engagement and high performance were, “The course was not available in my school.” Meanwhile, the majority of responses from students in courses that were categorized as low engagement and low performance were, “I was making up a class I had failed.” The level of engagement in our study may have been influenced by motivation. In addition, Lim & Morris (2009) found older students had a better chance to be successful in online learning at the higher education level. However, our study found older students were more likely to be at-risk students in K-12 online education. Students older than 18 tended to be low engaged and lower performing in their courses.

**Satisfaction and performance**

There is no confirmed relationship between student performance and satisfaction (positive correlation—Eiszler 2002, Nasser & Hagtvet 2006; No relation—Ladebo, 2003, Walker & Palmer, 2011). In the case study, students with higher final grades usually had higher course and instructor satisfaction. In addition, survey results showed female students reported significantly higher satisfaction with courses and instructors. Similar results were also found in Hermans, Haytko, & Mott-Stenerson’s study (Hermans, Haytko, & Mott-Stenerson, 2009).

Survey results showed that Health courses required the least effort and were considered the least challenging subject. Students also had high satisfaction with Health courses. However, although students had high satisfaction and engagement levels in Science courses, their average final grades were the third lowest. These findings showed high satisfaction and engagement levels could not guarantee high performance. While students obtained the lowest average grades in Math and English courses, they did not show significantly lower satisfaction levels in these two subject areas. Assuming perception data only reflected positive experiences, the picture of students’ experiences in courses could be misrepresented if students’ learning behaviors were not analyzed. These findings illustrate the flaws in solely relying on self-reporting and perception data in program evaluation to inform strategic decisions.

**Limitations**

It should be noted that this project was not without challenges. We are in the early stages of data mining, and learning management system providers are often not cognizant of the importance of their data collection strategies. For example, the Blackboard activity accumulator grouped wide-ranging learning behaviors into only five categories, which significantly simplified findings that could potentially be revealed by data mining. In addition, issues with missing data, data stored in the wrong fields, or mismatched data fields in the database limited some of the analysis techniques attempted in this evaluation. Blackboard failed to track ever forum reply behavior, and in order to avoid using inaccurate data to generate false results, we discarded students-student and student-instructor interaction analysis. Although multiple data sources provide rich information for data interpretation, some unexpected results
were revealed by survey design and investigation. For example, why did students from different schools show
different satisfaction levels for their instructors? Further data interpretation could be undertaken with additional
open-ended questions.

Proof of the framework

This study demonstrates the benefits of incorporating data mining into the program evaluation of K-12 online
education. It is important to note that demographic data and course evaluation survey data are also indispensable in
supporting data-mining-results evaluation and interpretation. In exploring EDM applications at the K-12 level that
have already been broadly adopted in higher education institutions we have demonstrated: 1) how data mining can be
incorporated into program evaluation in order to support decision-making at the institutional level, 2) how a
framework of data triangulation generates high-quality and non-partial results that can be combined with student
learning logs, demographic data and course evaluation surveys and that 3) characteristics of successful and at-risk
students can be generated by identifying important predictors of student performance, course satisfaction, and
instructor satisfaction for K–12 online education.

Future research

Future study should continue in the following directions: 1) Evaluate the usefulness of the information by surveying
administrators; and 2) Generalize findings by incorporating more data from various program evaluation projects. For
example, many average numbers in this study could be important indicators for retention and performance
improvement. The numbers would be more stable for generalization with additional study data; 3) Develop an early
warning system using tables and figures in this study to simplify administrators’ decision-making processes; 4) Add
further possible open-ended questions for data interpretation; 5) Conduct a follow up study. For example, entry-level
Science courses were identified as classes that might have course design issues. Follow up research might help
identify which parts are difficult for students; and 6) Utilize the framework to validate or generate educational
models.

References

Higher Education, 44(6), 657–672.

In B. V. Dasarathy (Ed.), Proceedings of Data Mining and Knowledge Discovery: Theory, Tools, and Technology II, 4057, 412–
419.


Chapman and Hall.


Education, 26(6), 525–38.

Education, 7(1), 31–54.

501.


Grammatikopoulos, V. (2012). Integrating program theory and systems-based procedures in program evaluation: A dynamic


Translating Learning into Numbers: A Generic Framework for Learning Analytics

Wolfgang Greller* and Hendrik Drachsler
Open University of the Netherlands, Valkenburgerweg 177, 6419AT Heerlen, The Netherlands // wolfgang.greller@ou.nl // hendrik.drachsler@ou.nl

*Correspondence author

ABSTRACT
With the increase in available educational data, it is expected that Learning Analytics will become a powerful means to inform and support learners, teachers and their institutions in better understanding and predicting personal learning needs and performance. However, the processes and requirements behind the beneficial application of Learning and Knowledge Analytics as well as the consequences for learning and teaching are still far from being understood. In this paper, we explore the key dimensions of Learning Analytics (LA), the critical problem zones, and some potential dangers to the beneficial exploitation of educational data. We propose and discuss a generic design framework that can act as a useful guide for setting up Learning Analytics services in support of educational practice and learner guidance, in quality assurance, curriculum development, and in improving teacher effectiveness and efficiency.

Furthermore, the presented article intends to inform about soft barriers and limitations of Learning Analytics. We identify the required skills and competences that make meaningful use of Learning Analytics data possible to overcome gaps in interpretation literacy among educational stakeholders. We also discuss privacy and ethical issues and suggest ways in which these issues can be addressed through policy guidelines and best practice examples.

Keywords
Learning analytics, Framework, Educational data mining, Ethics, Domain design, Data for learning

Introduction
In the last few years, the amount of data that is published and made publicly available on the web has exploded. This includes governmental data, Web2.0 data from a plethora of social platforms (Twitter, Flickr, YouTube, etc.), and data produced by various sensors such as GPS location data from mobile devices. In the wake of this, data-driven companies like Google, Yahoo, Facebook, Amazon, etc. are growing exponentially by commercially exploiting such data for marketing or in the creation of new personalised services. The new “data economy” empowers companies to offer an increasing amount of data products at little or no cost to their users (e.g., Google Flu Trends, bit.ly customised URLs, Yahoo Pipes, Gapminder.com). This growth in data also renewed the interest in information retrieval technologies. Such technologies are used to analyse data and offer personalised data products customised to the needs and the context of individual users.

It is already evident that data in combination with information retrieval technologies are not only the basis for the emergent data economy, but also hold substantial promises for use in education (Retalis et al., 2006; Johnson et al., 2011). One example of this is the research on personalisation with information retrieval technologies which has been a focus in the educational field for some time now (Manouselis et al., 2010). The main driver is the vision of improved quality, effectiveness, and efficiency of the learning processes. It is expected that personalised learning has the potential to reduce delivery costs while at the same time creating more effective learning experiences, accelerating competence development, and increasing collaboration between learners.

Not so long ago, for universities and companies alike, gathering data on their users met with substantial limitations in terms of cost, time requirements, scope, and authenticity of the data, as this was typically done using questionnaires or interviews with a selected representative number of stakeholders. The new data economy has made data collection very much an affordable activity. It is based on the highly economic electronic data mining of people’s digital footprints and the automated analysis of behaviours of the entire constituency rather than sampling. Because data mining is not a separate act to normal user behaviour, the information retrieved is also highly authentic in terms of reflecting real and uninterrupted user behaviour. As such, data mining is more comparable to observational data gathering than to intrusive collection via direct methods. This will not make questionnaires and structured interviews
obsolete, but it will greatly enhance our understanding and highlight possible inconsistencies between user behaviour and user perception (Savage and Burrows, 2007).

The proliferation of interactive learning environments, learning management systems (LMS), intelligent tutoring systems, e-portfolio systems, and personal learning environments (PLE) in all sectors of education produces vast amounts of tracking data. But, although these e-learning environments store user data automatically, exploitation of the data for learning and teaching is still very limited. These educational datasets offer unused opportunities for the evaluation of learning theories, learner feedback and support, early warning systems, learning technology, and the development of future learning applications. This leads to the importance of Learning Analytics (LA) being increasingly recognised by governments, educators, funding agencies, research institutes, and software providers.

The renewed interest in data science and information retrieval technologies such as educational data mining, machine learning, collaborative filtering, or latent semantic analysis in Technology-Enhanced Learning (TEL) reveals itself through an increasing number of scientific conferences, workshops and projects combined under the new research term Learning Analytics. Examples are the 1st Learning Analytics conference in Banff, Canada, 2011; the 4th International Conference on Educational Data Mining 2011 in Eindhoven, Netherlands; the 1st dataTEL workshop on Educational Datasets for Technology-Enhanced Learning at the Alpine-Rendez-Vous conference La Clusaz, France 2011; the 2nd International Conference on Learning Analytics and Knowledge (LAK12), Vancouver 2012; the 1st Workshop on Learning Analytics and Linked Data (LALD 2012); and more. Thus, the increasing amount of dedicated research events and publications make a meta-analysis of the domain timely and needed in order to establish a solid scientific basis which facilitates the development of new learner-oriented services.

**Critical dimensions of learning analytics**

Despite the great enthusiasm that is currently surrounding LA, it also raises substantial questions for research. In addition to technically-focused research questions such as the compatibility of educational datasets, or the comparability and adequacy of algorithmic and technological approaches, there remain several “softer” issues and problem areas that influence the acceptance and the impact of Learning Analytics. Among these are questions of data ownership and openness, ethical use and dangers of abuse, and the demand for new key competences to interpret and act on LA results. We shall point at these issues in more detail below. This means that the implementation of LA in learning processes requires to be carefully crafted in order to be successful and beneficial.

This necessity motivated us to identify six critical dimensions (soft and hard) of LA, which need to be covered by the design to ensure an appropriate exploitation of LA in an educationally beneficial way. By soft issues we mean challenges that depend on assumptions being made about humans or the society in general, e.g., competences or ethics. They are opposed by the hard challenges of the fact-based world of data and algorithms (cf. also the similar soft-hard distinction in Dron, 2011). In its coverage of soft issues, our framework differs from other, more workflow oriented models for LA, like that by Siemens (2011), although in his presentation he does acknowledge these as of concern. Rather than being a process model such as those collected in Elias (2011), we aim at a description framework that can later be developed into a domain model or ontology.

The critical dimensions highlighted here have been deduced from discussions in the emerging research community using a general morphological analysis (GMA) approach (cf. Ritchey, 2011). In this early formation stage of the LA community, scientific exchanges such as open online courses (MOOC) in Learning and Knowledge Analytics (LAK11, LAK12), or the above-mentioned events and congregations soon began to revolve around a number of key questions, like: Who is the target group for LA? What are we trying to achieve? How do we deal with privacy and data protection? These questions are naturally extended by other on-going debates such as the openness of data, which has been a topic for some time in the EDM and Open Linked Data domain, as well as technical and theoretical questions on achieving meaningful extraction of information from data.

Our chosen approach leading to the proposed framework consisted of a number of gathering and analysis processes. First, as a matter of opinion mining, we scanned the scientific interactions from proceedings and presentations of the conferences and working groups mentioned above. We conducted a brief literature review of abstracts in the field of Learning Analytics and Educational Data Mining. Additionally, we scanned the live discussions on the LA Google Groups (http://groups.google.com/group/learninganalytics and http://groups.google.com/group/LAK11), as well as
the LAK11 MOOC (presentation chats and social networking exchanges). Furthermore, we looked back at recent RTD projects that contained elements of analytics and the questions and lessons they produced, e.g., the Language Technologies for Lifelong Learning project (http://www.ltfll-project.org) contained an analytics approach related to learner positioning and conceptualisation. Following these reviews, we applied cognitive mapping (Ackermann, Eden, and Cropper, 2004) for synthesising and sense making. We analysed these discussions and clustered them into the proposed six fields of attention, which we then presented as the first draft of the framework to a community of commercial and academic experts for evaluation and feedback at the SURF seminar on Learning Analytics (Eindhoven, 30-31 August 2011). The number six in the framework is not chosen for any particular reason, and other divisions are of course possible. However, we find the dissection into these six dimensions a useful and easy to follow domain orientation.

With the framework, we take the presumption that responsible designers of analytic processes will not only implement what is technically possible to do and legally allowed (or at least not prohibited), but to consider holistically the outcomes for stakeholders and, even more importantly, the consequences for the data subjects, i.e., the people supplying the data (cf. the section on stakeholders below). The framework intends to be a guide as much as a descriptor of the problem zones. Hence we refer to it as a “design framework” that can and should be used to design LA services from an inclusive perspective. We will argue below that this will help the transferability of LA approaches between different contexts of application and research.

Proposed design framework for learning analytics

Our proposed model for the domain and application of LA in figure 1 below considers six critical dimensions. Each of the dimensions can be subdivided into several instantiations falling into that dimension. For example, the generic “stakeholder” dimension can have instantiations (values) like “learners” and “teachers.” The list of instantiations in the diagram is not exhaustive and can be extended on a case-by-case basis. To stay with the above example, commercial service providers and even automated agents could also function as stakeholders in a LA process. It is useful to note that through connecting various (and also multiple) different instantiations of each dimension, concrete use cases can be constructed. We call the dimensions “critical” in the sense that each of the six fields of attention is required to have at least one instantiation present in a fully formulated LA design. We realise, though, that some dimensions are vaguer than others in this respect.

![Diagram of the proposed design framework for learning analytics](image)

Figure 1. Critical dimensions of learning analytics
The six dimensions of the proposed LA framework are (cf. Figure 1): stakeholders, objectives, data, instruments, external constraints, and internal limitations. We will discuss each of these dimensions individually in the following and exemplify their instantiations and impact on the LA process and the benefits and opportunities they may determine. We will also elaborate apparent problem zones and limitations that may hinder any envisaged benefits.

Before embarking on the abstract dimensions in detail, we would like to illustrate the purpose and possible usage of the framework on the following sample use case, which is created out of a number of instantiations of the six dimensions. This specific example relates to conducting a social network analysis of students discussing in a forum using the SNAPP tool, based on the work by Dawson et al. (Dawson, 2008; Macfadyen & Dawson, 2010).

<table>
<thead>
<tr>
<th>Table 1. Sample use case and values for dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimension</strong></td>
</tr>
</tbody>
</table>
| Stakeholders | *Data subjects:* a group of learners.  
*Data clients:* tutor, discussion moderator. |
| Objective | *Reflection:* Analyse student interactions in a forum discussion, identify network connections between students, and identify isolated students to bring them back into the discussion. |
| Data | *Protected dataset:* Student interactions and posts in the discussion forum of the LMS.  
*Relevant indicators:* Posts published, posts replied to.  
*Time scale:* what time frame is applied to the analysis? |
| Instruments | *Pedagogic theory:* socio-constructivist, hypothesis is that active participants in a discussion show better learning outcomes.  
*Technology:* Social Network Analysis (SNA), statistics.  
*Presentation:* network diagram visualisation, stats table. |
| External limitations | *Conventions:* (1) Privacy: is the analysis in accordance with privacy arrangements, are the students properly informed?  
(2) Ethics: What are the dangers of abuse/misguided use of the data?  
*Norms:* Are there e.g., legal data protection or IPR issues related to this kind of use of student data?  
*Time scale:* will the students still be able to benefit from the analytics outcome? Is the analysis post-hoc or just-in-time? |
| Internal limitations | *Required competences:* (1) Interpretation: Do the data clients have the necessary competences to interpret and act upon the results? Do they understand the visualisation or presentation of the information? (2) Critical thinking: Do they understand which data is represented and which data is absent? How will they use this information? |

The above use case can be used (1) as a checklist when designing a purposeful LA process; (2) as a sharable description framework to compare context parameters with other similar approaches in other contexts, or for replication of the scientific environment. The framework allows an indefinite number of use cases with the respective value arguments.

**Stakeholders**

The stakeholder dimension includes *data clients* as well as *data subjects*. Data clients are the beneficiaries of the LA process who are entitled and meant to act upon the outcome (e.g., teachers). Conversely, the data subjects are the suppliers of data, normally through their browsing and interaction behaviour (e.g., learners). It is important to make this distinction in order to understand the impact of the process on individuals. In certain cases, the two types of stakeholder groups can be the same, as is the case if a LA application feeds back information to learners about their own learning rather than to inform the teacher, as would be a common case in informal learning scenarios. In the traditional learner-teacher scenario, the teacher would act as the data client, who receives information gathered from the data subjects, i.e., the learners.
As shown in the framework model (Figure 1), the main stakeholder groups of LA in formal learning situations are learners, teachers, and educational institutions. These may be expanded or substituted by other stakeholder groups, such as researchers, service providers, or governmental agencies. Each of the groups has different information needs and can be provided with tailored views on information using LA.

Information flow between stakeholders can best be exemplified with the common hierarchical model taken from formal education (Figure 2). What the diagram illustrates as an example is by which ways benefits might be obtained from LA. The pyramid encapsulates the academic layers of education and training institutions. In the most direct way, data analysis from the student level, e.g., via a LMS, can inform the above layer, in this case the teachers. Teachers can then use the analytics information to plan targeted interventions or adjust their pedagogic strategies. Institutions can, similarly, retrieve benefits from student and teacher data in order to provide staff development opportunities or to plan policies like quality assurance and efficiency measures. We also want to stress the major benefits LA offers for self-reflection on every level (cf. left side of the diagram). We would like to see institutions enabling and actively encouraging students to reflect on their learning data. But also teachers and institutions can gain new insights by reflecting on their performance. Not immediately involved in the learning processes, researchers (right of the diagram) could harvest data for the purpose of evaluating or innovating teaching processes or learning services. Finally (on top of the diagram), Government agencies may collect cross-institutional data to assess the requirements of Higher Education Institutes (HEI) and their constituencies.

![Figure 2. Information flow between LA stakeholders](image)

Although they are the most widespread form in formal education, hierarchies are not the only flow models to describe where benefits can be retrieved. For example, peer evaluation using Personal Learning Environments (PLE) may be another information environment for LA. Peer environments also prevail in academic transactions like conferences or publications that are based on peer review systems. Practical examples for a horizontal peer-related information flow are the various scientific impact measures that exist, e.g., citation indexes. Equally, serious games can provide a non-hierarchical approach and/or team perspective to collaborative learning, e.g., how fast a team completed a level. In each of these, however, lie some issues of dependency and possible legal constraints (cf. further below).

Example opportunities for LA with respect to different stakeholder groups are:

Students can be supported with specific learning process and reflection visualisations that compare their performance to the overall performance of a course. Furthermore, they can be provided with personalised recommendations for suitable learning resources, learning paths, or peer students (Gaviria et al., 2011).
Teachers can be provided with course monitoring systems that inform them about knowledge gaps of particular pupils and thus enable them to focus their attention on those pupils. They can also harvest emergent group models that can lead to shared understanding of domain topics or processes for better curriculum design and on-the-fly adaptations.

Institutions can monitor the performance of students regarding drop-out and graduation rate on a much finer granular level. In this way, they can evaluate their courses and improve outcomes of their courses.

Other stakeholders: We would like to emphasise that stakeholders need not be confined to formal education settings, but include all formal, non-formal, or informal environments, such as professional development (CPD). In these cases, the stakeholders are to be substituted by the relevant entities. For non-formal learning, for example, stakeholders would include a “learner” instantiation with (only) a self-reflection dimension in which feedback is mirrored back to the same person. In work-based learning, employees and line-managers may be the most common stakeholder groups involved. More notably, computer agents can also serve as stakeholders, for example as data clients that take further decisions on the learner’s behalf or trigger an event (e.g., notification e-mail, recommendation of content or peer, etc.).

Objectives

The main opportunities for LA as a domain are to unveil and contextualise so far hidden information out of the educational data and prepare it for the different stakeholders (see above). Monitoring and comparing information flows and social interactions can offer new insights for learners as well as improve organisational effectiveness and efficiency. This new kind of information can support individual learning processes but also organisational knowledge management processes (Butler & Winne, 1995). We can distinguish two fundamentally different objectives: reflection and prediction (cf. Figure 1 above).

Reflection: Reflection is seen here as the critical self-evaluation of a data client as indicated by their own datasets in order to obtain self-knowledge. Wolf (2009) calls this process the “quantified self”, i.e., self-observation and reacting to one’s own performance log data. There already is a growing number of Personal Informatics Systems, i.e., human-computer interaction systems that support this process (Li & Forlizzi, 2010). However, reflection may also be seen as critical self-evaluation based on other stakeholders’ datasets. This would especially be true if, for example, a teacher was led to reflect upon their teaching style as indicated by the datasets of their students. In the above hierarchical flow model (Figure 2), the higher order stakeholder would have the ability to utilise all the datasets from lower constituencies for their own reflection.

On an individual level, LA can support reflection of learning processes and offer personalised information on the progress of the learner (Govaerts et al., 2010). On the institutional level, LA can enhance monitoring processes and suggest interventions or activities for particular students. Greatest care should however be taken not to confuse objectives and stakeholders in the design of a LA process and not to let, e.g., economic and efficiency considerations on the institutional level dictate pedagogic strategies, as this would possibly lead to industrialisation rather than personalisation.

LA is a support technology for decision making processes. Therein also lies one of the greatest potential dangers. Using statistical analytic findings is a quantitative not a qualitative support agent to such decision making. We are aware that aligning and regulating performance and behaviour of individual teachers or learners against a statistical norm without investigating the reasons for their divergence may strongly stifle innovation, individuality, creativity and experimentation that are so important in driving learning and teaching developments and institutional growth.

Prediction: Apart from support for reflective practice, LA can also be used for predicting and modelling learner activities (Siemens, 2011; Verbert et al., 2011). This can lead to earlier intervention (e.g., to prevent drop-out), or to adaptive services and curricula. Using Machine Learning techniques, for example, learner profiles can be built dynamically and automatically, saving the learner filling in and maintaining profile data. In predictive outcomes lies currently much hope for efficiency gains in terms of establishing acts of automatic decision making for learning paths, thus saving teacher time for other more personal interventions. But prediction suffers potentially from big ethical problems (to which more further below), in that judgements about a person, whether originating from another
human or a machine agent, if based on a limited set of parameters could potentially limit a learner’s potential. For example, not every learner who has difficulties mastering subject level two, will automatically not master level three. We have to prevent re-confirming old-established prejudices of race, social class, gender, or other with statistical data, leading to restrictions being placed upon individual learners. Furthermore, there are limitations in the use of LA data as a means for supporting the learning process. Learning processes assume the leading role of the learner, rather than that of the teacher. However, the reliability of a LA-supported learner profile and its usefulness to the learners will remain questionable. For example, what LA data can be used in order to define whether a learning activity had a “high” or “low” impact on the learning process of learners, and at which points in the process itself? The diversity of learning makes it also problematic to judge which learning activity was of high value for learner A but of low value for learner B.

With respect to pedagogic theories, we would like to argue that LA does neither support nor ignore specific pedagogic theories, and as an abstract concept is pedagogically neutral. Indeed, we are of the opinion that LA can be used to evaluate different pedagogic strategies and their effects on learning and teaching through the analysis of learner data. This can be defined as a specific pedagogically oriented objective under the current dimension, but, as we will discuss further below, certain technologies are not pedagogically neutral and this will influence the analytics process in one way or another.

**Educational data**

LA takes advantage of available educational datasets from different Learning Management (LMS) and other systems. Institutions already possess a large amount of student data, and use these for different purposes, among which administering student progress and reporting to receive funding from the public authorities are the most commonly known. Linking such available datasets would facilitate the development of mash-up applications that can lead to more learner-oriented services and therefore improved personalisation.

LA strongly relies on data about learners and one of the major challenges LA researchers are facing is the availability of publicly available datasets to evaluate their LA methods. Most of the data produced in institutions is protected, and the protection of student data and created learning artefacts is a high priority for IT services departments. Nevertheless, similar to Open Access publishing and related movements, calls for more openness of educational datasets have already been brought forward (Drachsler et al., 2010). Anonymisation is one means of creating access to so-called Open Data. Recently, Verbert et al., (in press) presented a state of the art review of existing educational datasets. How open educational data should be, requires a wider debate (cf. section on legal constraints below), but, already in 2010, several data initiatives where started to make more educational data publicly available:

**dataTEL challenge**—The first dataTEL challenge was launched as part of the first workshop on Recommender Systems for TEL (Manouselis et al., 2010), jointly organized by the 4th ACM Conference on Recommender Systems and the 5th European Conference on Technology Enhanced Learning (EC-TEL 2010) in September 2010. In this call, research groups were invited to submit existing datasets from TEL applications that can be used for LA research purposes and recommender systems for TEL.

**dataTEL workshop**—The “Datasets for Technology Enhanced Learning” workshop was organised at the third STELLAR Alpine Rendez-Vous in March 2011. During this workshop, related initiatives that are collecting educational datasets, and apply these in data-driven learning applications were presented, and challenges related to privacy and data protection were discussed.

**PSLC dataShop** (Stamper, 2011) offers an open data repository that provides access to a large number of educational datasets. *dataShop* has data from students derived from interactions with intelligent tutoring systems.

**LinkedEducation.org** (Dietze et al., 2012) is another initiative that provides an open platform to promote the use of data for educational purposes. At the time of writing, five organizations have contributed datasets.
Despite these pioneering activities, it does, by comparison, still seem somewhat bizarre that in the commercial world with clicking the “register” button, the default access to all user data becomes owned by some company, whereas educational institutions operate on the default that everything is protected from virtually everyone.

Distinguishing educational data by access rights in open and protected datasets (Figure 1) is not as simple as it sounds. Because the technical systems producing and collecting data are typically owned by the institution, the easiest assumption would be that this data belongs to them. However, which employees of the institution exactly are included in the data contract between a learner (or their parents) and the educational establishment, is as yet unresolved. This poses severe constraints on inner-institutional research or wider institutional use. We will bring up some more legal consideration under the point on external constraints below.

Like in related research domains, LA datasets create a new set of challenges for research and practice. These include:
- A lack of common dataset formats like the suggested one from the CEN/ISSS PT social data group (cf. CAM Schema at: https://sites.google.com/site/camschema/home; and Wolpers et al., 2007).
- The need for version control and a common reference system to distinguish and point to different datasets.
- Methods to anonymise and pre-process data according to privacy and legal protection rights (Drachsler et al., 2010).
- A standardised documentation of datasets so that others can make proper use of it like that promoted by the data-seal-of-approval initiative (cf. http://www.datasealofapproval.org).
- Data policies (licences) that regulate how users can use and share certain datasets. For instance, the Creative Commons licensing rights could be considered as a standard way to grant permissions to datasets. DataCite (Brase, 2009) is an organization that enables to register research datasets and to assign licensing rights to them, so that the datasets can be referenced similar to academic articles.

From a technical point of view, idealised datasets probably remain the biggest challenge for analytics. This is to say that the assumption that datasets consist of context-free, meaningful and only meaningful data, is highly optimistic. In most natural settings, users “pollute” databases by producing erroneous or incomplete datasets. For example, teachers who want to see their students’ view on LMS courses often set themselves up as “test students” or create “test courses”. These are not always obvious, but need to be removed from the data to be analysed. Therefore empirical findings coming from a specific dataset are almost certainly affected by the context of data collection and processing.

Similarly, data collection often leads to “enmeshed identities” being used for analytics and prediction. A dataset cannot typically distinguish between a single individual and a shared presence in the learning space (group work on a single device). Students who often work together with others on shared devices (laptops, smartphone, lab space, etc.) produce enmeshed fingerprints in their educational data. This may lead to behaviours being attributed to a logged-in identity that may actually have originated from an “invisible” partner. Standardised documentation of datasets can be seen as paramount to raise awareness of this danger.

Additionally, from a pedagogic perspective, it remains an on-going challenge to formulate indicators from the available datasets that bear relevance for the evaluation of the learning process. The selection of specific data and their weighting (under the methods applied in the “instruments” dimension) against the real behaviour of students is of greatest importance, as is the process of relating behaviour pattern data to cognitive developments.

**Instruments**

Different technologies can be applied in the development of educational services and applications that support the objectives of educational stakeholders. LA takes advantage of so-called information retrieval technologies like educational data mining (EDM; cf. Romero et al., 2008), machine learning, or classical statistical analysis techniques (cf. Figure 1), but other techniques may also be considered relevant, e.g., social network analysis (cf. Buckingham & Ferguson, 2011) or Natural Language Processing (NLP).

Through these technologies, LA can contribute tailored information support systems to the stakeholders and report on demand. For instance, LA could be applied to develop a drop-out alert system. High drop-out rates are a challenging problem in education, especially distance education. Further research on LA can contribute to decrease
the drop-out rate by developing e.g., a *Drop-out Analyser* that notifies the teacher of a course in time which students are in danger of falling behind or dropping out. This could be done by using LMS datasets and train a certain information retrieval technology (e.g., a Bayesian classifier) on the datasets to learn behavioural patterns of students that dropped out. Afterwards, the *Drop-out Analyser* could be applied on a follow-up online course and flag up students that show similar patterns. The teacher of the course could then intervene in an appropriate manner. Preliminary prototypes of such systems are already available, like the Blackboard Early Warning System.

Under the dimension “instruments” in our model (Figure 1), we also subsume conceptual instruments such as theoretical constructs, algorithms, or weightings, by which we mean different ways of approaching data. These ways in the broadest sense “translate” raw data into information. The quality of the output information and its usefulness to the stakeholders depend heavily on the methods chosen. Hildebrandt (2010), quite rightly, warns that “invisible biases, based on … assumptions … are inevitably embodied in the algorithms that generate the patterns”.

Competing methods, technologies and algorithms applied to the same set of data, will result in different outcomes, and thus may lead to different consequences in terms of decision making based on these outcomes. LA designers and developers need to be aware that any algorithm or method they apply is reductive by nature in that it simplifies reality to a manageable set of variables (cf. Verbert et al., 2011).

**External constraints**

Many different kinds of constraints can limit the beneficial application of LA processes, some being “softer” than others. It has been suggested to us to identify them as ethical, legal, and social constraints, but also to feature organisational, managerial, and process constraints. This we find a useful subdivision of external limitations, but other divisions look equally logical. In the abstraction of the diagram above (cf. Figure 1), we propose the preliminary distinction of *conventions*, under which we count ethics, personal privacy, and similar socially motivated limitations, and, *norms* that are restricted by laws or specific mandated policies or standards. For reasons of space, we want to elaborate especially on the ethical aspects as this has grown into a field of much recent attention and debate (Bollier, 2010) and even spawned a collaborative effort in the Learning Analytics research community (Siemens, 2012).

New ethical and privacy issues arise when applying LA in education. These are challenging and highly sensitive topics when talking about educational datasets, as described in Drachsler et al. (2010). The feeling of endangered privacy may lead to resistance from data subjects toward new developments in LA. In order to use data in the context of LA in an acceptable and compliant way, policies and guidelines need to be developed that protect the data from abuse. Legal data protection may require that data is anonymised before it can be used. At the same time, as much openness of the datasets as possible is desirable (see paragraph on data above).

Personal data enjoys strong legal protection, differing by national laws and sometimes competing with other legal frameworks such as the Freedom of Information Act in the United Kingdom. We will not go into the legal details here, but did already above hint at the predicament that faces any data-related venture when using people’s digital footprints. It concerns the lack of legal clarity with respect to data ownership. In current circumstances, data gathered about a person (before it is anonymised) belongs to the owner of the data collection tool, typically also the data client and beneficiary. Up till now, through direct intervention like questionnaires and sign-up processes, this was not a big problem to the data subjects. However, with the dramatic increase of ambient sensors and new technologies, such as location tracking or biometric face-recognition cameras, etc., more and more parts of individual behaviour are logged without the data subject’s approval or even awareness. That being so, the ethic principle of “informed consent” (cf. AoIR, 2002) is very much under threat. The fundamental question legislators need to ask is: who does a person’s life data belong to? (cf. also Hildebrandt, 2006). We believe that this question may in the near future become more and more elevated in importance and prominence.

On institutional level, educational and student data was traditionally handled separately, and is legally something of a blind spot. Registration data was kept and maintained by registry staff, IT data by IT staff, and learning data by academic staff. To use LA to its full potential, integration of available institutional datasets needs to happen. Universities, for instance, already collect and report socio-economic data such as students’ post codes or ethnic and linguistic background. Institutions are even legally obliged by funding bodies to do so, but integrating this dataset...
with educational performance data, would be widely considered unethical or even illegal. As has been already mentioned above, the extent of a student’s data contract with an institution and its individual staff representatives in different roles (teacher, administrator, secretary, researcher, IT support staff, Deans and management, etc.) needs to be urgently clarified. At the same time, privacy directives such as the Data Seal of Approval supported by the Dutch DANS institute (http://www.datasealofapproval.org) and related European data directives like the European Directive on data protection 95/46/EC (Directive, EU, 1995) need to be implemented.

Even where in compliance with the law, educational data can easily be abused for purposes inappropriate for educational institutions or for the data subjects (especially where minors are concerned). By principle, the more access to information about a data subject a data client has, the higher the responsibility is to use this information in a sensitive and ethical way. In an inspired article, Hildebrandt (2010) elaborates the ethically limited applicability of automated pattern recognition to the Law domain, but these limitations can be transferred just as easily to the domain of learning. Among the more obvious ethical risks are the exploitation of such data for commercial and similar purposes, or data surveillance issues (social sorting, cumulative disadvantages, digital stalking) and their ethical implications.

Ethics don’t stop at the data gathering and integration. The realisation that we may encounter conflicts in values and interests in and through the analysis of people’s behaviour needs to guide the post-analytic decision making process and the conclusions drawn from the approach. It is important to remind stakeholders of LA processes that data can be interpreted in many ways and lead to very different consequent actions. To give a drastic example, imagine being confronted with the insight that children from an immigrant background show reading difficulties, backed by supportive data analysis. This may lead to a wide ranging variety of responses, from developing extracurricular support mechanisms, to segregated classes, up to bluntly racist abuse of various kinds.

Ethics in LA may affect students and teachers alike, especially, where institutions aim to use LA to quality assure the performance of their teaching staff. Data can easily be abused as supporting evidence for exercising inappropriate pressures on data subjects to change otherwise perfectly acceptable or explainable performance behaviour. Institutions are therefore challenged with establishing a set of ethical policies and principles, together with, e.g., complaints procedures and safety nets that secure proper use of educational data in teaching and research. We find the ethical guidelines of the Association of Internet Researchers, AoIR (2002), a useful starting point in this respect, in that it has a purpose-oriented approach that supports ethical pluralism and respects the individual.

Another ethical consideration is the acceptance of divergence in the data constituency (AoIR, 2002). We already touched upon the danger that the result of algorithmic analysis, consequent policies and exercised pressures may aim at uniformity and at mainstreaming learning and teaching processes, thereby greatly harming creative processes and innovation that diverge from the statistical mean. It is one of the principal shortfalls of statistical prediction that it can only predict average behaviour not outliers. As such, LA provides no means of predicting exceptions to a rule, or exceptions to the exception rule.

**Internal limitations**

In complement to the environmental problem areas contained in the above “external constraints” section, we subsume a number of human factors that enable or may pose obstacles and barriers under the dimension heading “internal limitations.” Prominent among these are competences and acceptance.

It is already becoming clear that the application of learning analytics requires new higher-order competences to enable fruitful exploitation in learning and teaching. In order to make LA an effective tool for educational practice, it is important to recognise that LA ends with the presentation of algorithmically attained results that require interpretation (Reffay & Chanier, 2003; Mazza & Milani, 2005). There are innumerable ways to present and to interpret data and base consequent decisions and actions on it, but only some of them will lead to benefits and to improved learning. Basic numeric and other literacies, as well as ethical understanding are not enough to realise the benefits that LA has to offer. In a recent survey we conducted among LA experts, only 21% of the 111 respondents felt that learners would possess the required competences to interpret LA results themselves and determine appropriate actions/interventions from it (Drachsler & Greller, 2012). Therefore, the optimal exploitation of LA data requires some high level competences in this direction, but interpretative and critical evaluation skills (cf. Figure 1)
are to-date not a standard competence for the stakeholders, whence it may remain unclear to them what to do as a consequence of a LA outcome or visualisation.

Interpretation of LA results is often facilitated by enticing visualisations that are aimed to serve as a functional aid (Figure 3). One inherent danger that we perceive is that the simplicity and attractive display of data information may delude the data clients, e.g., teachers, away from the full pedagogic reality. This may negatively affect the pedagogic assessment and grading of a student’s performance which should not be based alone on the visualisation of log files from a Learning Management System. To illustrate this danger take the example of marking student essays. An automated spell-check on orthographic mistakes presents itself as a quick and simple to interpret translation of the learner artefact into numbers. This makes it ideal for an efficient, cognitively effortless, and egalitarian grading mechanism. In an education environment that increasingly suffers from time constraints and calls for more efficiency in teacher activities, it is easily imaginable that the traditional qualitative assessment of essays gives way to such quick number crunching being over-proportionally reflected in student marks.

In our model (cf. Figure 1), we include among the key competences for LA, critical evaluation skills, because superficial digestion of data presentations can lead to wrong conclusions. It has to be strongly emphasised that data not included in the respective LA approach, is equally if not more important than the dataset that is included. To judge a learner’s performance merely on, e.g., LMS quantitative data is like looking at a single puzzle piece. As learning is more and more happening in a lifelong and diverse ecosystem, an exclusive data view on single elements may provide a stimulus for reflection but not a sound basis for assessment.

The necessary competences notwithstanding, acceptance factors can further influence the application or decision making that follows an analytics process. This can, as is regularly seen in political debates, lead to blunt rejection of the results or applied methods from the constituency or parts thereof. In a learning context, ways to increase acceptance is vitally important also in order to produce usable outcomes. To get a better grasp on this issue, current scientific debate, therefore, should focus on empirical evaluation methods of learning analytics tools (Ali et al., 2012) and on advanced technology acceptance models (cf. Venkatesh & Bala 2008), inspired by the early work in this area (Davis, 1989, 1993). For LA, a revised technology acceptance model (TAM) could be an interesting approach to evaluate the emergent analytic technologies for all stakeholders described in our framework and also the needed implementation requirements to guarantee successful exploitation.

The place of pedagogy in the learning analytics framework

LA holds promises in the context of TEL by offering new methods and tools to diagnose learner needs and provide personalised instructions to better address these needs. It is not yet clear to what extent LA will lead to more personalised learning experiences rather than merely clustering people into behaviouristic “learner models” (e.g., as
“outliers” of mainstream models). Consequently, more empirical evidence is needed to identify which pedagogic theory LA serves best. LA has been effectively used for behaviourist-instructivist style approaches (but see the critical reflection by Pardo & Kloos, 2011), but there is as yet little evidence for the support of constructivist approaches to learning (Duffy & Cunningham, 2001), where learning is seen as an active cognitive process in which learners construct their own concepts of the world around them. In LA, the latter is mostly inferred indirectly, by relating grades of learning outcomes with activities during the learning process (Dawson, 2012). In these correlations, it emerges that active students get better results. However, the role LA plays in this has not yet been conclusively demonstrated. Despite these questions, we would like to maintain that as knowledge and experiences vary considerably among learners, the diversity of learning can more effectively be addressed by LA methods than with current learning environments.

In our model, LA can work in support of a multitude of pedagogic strategies and learning activities as manifested and represented by the available data. This means we can only see pedagogies through the data. Because of this, we do not include them as part of the analytics process (Figure 1) but as implicitly contained in the input datasets that encapsulate the pedagogic behaviour of users. As we know, this behaviour depends a great deal on the platform and the pedagogic vision the developers built in (Dron & Anderson, 2011). For example, data from a content sharing platform will have a behaviourist/cognitivist pedagogy attached to the learner behaviour, since this is the pedagogic model underlying the technology. In any case, only the pedagogic patterns exhibited in the dataset can be analysed; and this will vary.

Additionally, pedagogy can be explicitly addressed in the goals and objectives that the LA designer sets (“objectives” dimension). The LA method (“instruments” dimension) will determine the outcome of the analysis and together with the interpretation applied may lead to a large variety of options for consequences and interventions. If such pedagogic interventions were applied, they would lead to new behaviours which, once again, could be analysed through the available data (Figure 4). In the graph below, we refer to pedagogic behaviour as learner/teacher behaviour that is motivated by didactic designs (learning designs). Pedagogic consequences, similarly, are adjustments to the didactic strategy or learning design based on the outcomes of the LA process.

![Figure 4. Learning analytics and pedagogy](image)

A simple analogy would be boiling water in a pan. At any time (or continuously) you can immerse a thermometer and measure its temperature. The goal would be to determine whether you need to turn up the heat or not. The result of the analysis can then lead to the actions you want to take. The thermometer is only one method for such an analysis. An alternative would be to observe and wait until the water bubbles. Setting a benchmark (in the objectives design) can inform you when it is time for the teabag to go in. When applied to a learning process, immersing the thermometer into the water equates to the LA data gathering and analysis of learning in progress. It is here where learning is translated into numbers.

It should be noted that the pedagogic input factors are not confined to behaviour alone, but also include beliefs, (societal) values, and implicit theories of knowledge and learning. However, the LA application can only see these in the way they manifest themselves in the data. We also want to point at the possibility to apply computer agents to
determine specific interventions. These could be as simple as sending a notification or recommendation to one of the stakeholders.

An example for pedagogic consequences is the following (Dawson, 2012): Using SNAPP as a tool to do a social network analysis (SNA) on discussion forums in a learning environment, the moderator or tutor might discover that certain changes in the moderation, the organisation, or the task, may lead to more or less engaged discussion among participants. In this way, the information gained through the LA process can support the fine-tuning of pedagogic effectiveness in a particular activity, depending on the desired learning outcome.

The model takes note that pedagogic success and performance is not the only thing that LA can measure. LA collects snapshots taken from educational datasets. These snapshots can be used to reflect or predict, in order to make adjustments and inform interventions, either by a human or by a system. Apart from offering efficiency benchmarking and business information for education providers, new support services for learning and more qualitative personal experiences can be achieved.

**Conclusion and outlook**

In summary, the proposed framework model in figure 1 above stresses the inherent connections between the six different dimensions and the impact of the analytics process on the end user and the data suppliers. If one of the parameters changes, the outcome and anticipated benefits will change. It is therefore our conviction that only the consideration of all six dimensions in the design process can lead to optimal exploitation of LA. Additionally, substantial work on new ethical guidance, data curation, and ownership needs to happen at universities and in legislation to reduce the risks connected to the application of LA and to protect the data subject, usually the learner.

Because of the inherent dependencies, we argue that all six dimensions are mandatory to be argumentatively present in a fully flexed LA design. We would, therefore, strongly welcome if application developers and researchers would not only make their technical environment known and open, but also describe the contextual environment and expectations from the users (e.g., required competences) along the lines of the framework. This would allow scientific comparison, replication, external analysis, and alternative contextualisation.

To validate the framework as both a descriptive approach as well as a guide to the design process of LA applications, we suggest evaluating the growing number of LA application showcases and testing for consistency in the descriptive values of the model. Additionally, we want to create a selected number of use cases that encompass the six dimensions and their instantiations.

LA is very much at the dawn of its existence and considered by many as one of the technological advances that will bring learning onto the next higher level. While we join in with this chorus of positive expectations, we are also aware that LA shows facets of a double nature: In its most optimistic outlook, learners will be provided with personal information about their current needs, while, at the same time, the educational system will be evolved from a “one-size-fits-all” approach into a highly personal competence-driven educational experience. But this view is not without flaws, because of the real dangers that the extended and organised collection of learner data may not so much bring added benefits to the individual, but instead provides a tool for HEIs, companies, or governments to increase manipulative control over students, employees, and citizens, thereby abusing LA as a means to reinforce segregation, peer pressure, and conformism rather than to help construct a needs-driven learning society. We therefore believe that it will be of critical importance for its acceptance that the development of LA takes a bottom-up approach focused on the interests of the learners as the main driving force.

LA has the potential for new insights into learning processes by making hitherto invisible patterns in the educational data visible to researchers and end users, and to enable development of new instruments for everyday educational practice. However, there are substantial uncertainties about the extent of impact LA will have on education and learning in general. The proposed framework model is motivated by the potential and opportunities that LA offers in its relevance for educational development and opportunities to personalise learning. While we agree with the Horizon report’s forecast and its claim for a prosperous future of LA (Johnson et al., 2011), we also strongly feel that this development should not happen without a guiding framework that combines use of educational data with the protection of individuals and their learning.
Decisions based on LA are of concern, because they determine the usefulness and consequences for the stakeholders as well as the extent of its impact. Data analysis could have dramatic (and unwanted) consequences if not used with the necessary care. It is here where ethics play an enormously important role. Building of trust and confidence throughout the data constituencies has to be a priority from the start, and, here again, this proposed framework hopes to act as a useful guide.

One of the major questions in LA is the relation with theories of learning, teaching, cognition and knowledge. We hinted above at the opportunity that LA may support the evaluation of concrete didactic approaches which in turn may provide supportive evidence for particular pedagogic theories of learning and knowledge. At the same time, technologies are not pedagogically neutral; hence the evaluation will be influenced by the approach chosen. We consider this debate as an on-going one which will require further research and demonstration of applications and the impact they make on the process of learning.

It is still too early to base education fully on LA approaches alone, and we expect it never will be possible to do so. However, at the very least, opportunities this new discipline has to offer are to provide new support for learning activities and stimuli for reflection. In our opinion, it is these opportunities that LA should pursue.

References


Design and Implementation of a Learning Analytics Toolkit for Teachers

Anna Lea Dyckhoff, Dennis Zielke, Mareike Bültmann, Mohamed Amine Chatti and Ulrik Schroeder

Computer Supported Learning Group, RWTH Aachen University, D-52056 Aachen, Germany // dyckhoff@informatik.rwth-aachen.de // chatti@informatik.rwth-aachen.de // Schroeder@informatik.rwth-aachen.de // dennis.zielke@rwth-aachen.de // mareike.bueltmann@rwth-aachen.de

ABSTRACT

Learning Analytics can provide powerful tools for teachers in order to support them in the iterative process of improving the effectiveness of their courses and to collaterally enhance their students’ performance. In this paper, we present the theoretical background, design, implementation, and evaluation details of eLAT, a Learning Analytics Toolkit, which enables teachers to explore and correlate learning object usage, user properties, user behavior, as well as assessment results based on graphical indicators. The primary aim of the development of eLAT is to process large data sets in microseconds with regard to individual data analysis interests of teachers and data privacy issues, in order to help them to self-reflect on their technology-enhanced teaching and learning scenarios and to identify opportunities for interventions and improvements.

Keywords

Learning analytics, EDM, Improving educational software, Teacher support

Introduction

Learning Management Systems (LMS) or Virtual Learning Environments (VLE) are widely used and have become part of the common toolkits of educators (Schroeder, 2009). One of the main goals of the integration of traditional teaching methods with technology enhancements is the improvement of teaching and learning quality in large university courses with many students. But does utilizing a VLE automatically improve teaching and learning? In our experience, many teachers just upload existing files, like lecture slides, handouts and exercises, when starting to use a VLE. Thereby availability of learning resources is improved. For improving teaching and learning it could be helpful to create more motivating, challenging, and engaging learning materials and e.g., collaborative scenarios to improve learning among large groups of students. Teachers could e.g., use audio and video recordings of their lectures or provide interactive, demonstrative multimedia examples and quizzes. If they put effort in the design of such online learning activities, they need tools that help them observe the consequences of their actions and evaluate their teaching interventions. They need to have appropriate access to data to assess changing behaviors and performances of their students to estimate the level of improvement that has been achieved in the learning environment.

With the establishment of TEL, a new research field, called Learning Analytics, is emerging (Elias, 2011). This research field borrows and synthesizes techniques from different related fields, such as Educational Data Mining (EDM), Academic Analytics, Social Network Analysis or Business Intelligence (BI), to harness them for converting educational data into useful information and thereon to motivate actions, like self-reflecting ones previous teaching or learning activities, to foster improved teaching and learning. The main goal of BI is to turn enterprise data into useful information for management decision support. However, Learning Analytics, Academic Analytics, as well as EDM more specifically focus on tools and methods for exploring data coming from educational contexts. While Academic Analytics take a university-wide perspective, including also e.g., organizational and financial issues (Campbell & Oblinger, 2007), Learning Analytics as well as EDM focus specifically on data about teaching and learning.

Siemens (2010) defines Learning Analytics as “the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.” It can support teachers and students to take action based on the evaluation of educational data. However, the technology to deliver this potential is still very young and research on understanding the pedagogical usefulness of Learning Analytics is still in its infancy (Johnson et al., 2011b; Johnson et al., 2012).
It is a current goal at RWTH Aachen University to enhance its VLE—the learning and teaching portal L²P (Gebhardt et al., 2007)—with user-friendly tools for Learning Analytics, in order to equip their teachers and tutors with means to evaluate the effectiveness of TEL within their instructional design and courses offered. These teachers still face difficulties, deterring them from integrating cyclical reflective research activities, comparable to Action Research, into everyday practice. Action Research is characterized by a continuing effort to closely interlink, relate and confront action and reflection, to reflect upon one’s conscious and unconscious doings in order to develop one’s actions, and to act reflectively in order to develop one’s knowledge.“ (Altrichter et al., 2005, p. 6). A pre-eminent barrier is the additional workload, originating from tasks of collecting, integrating, and analyzing raw data from log files of their VLE (Altenbernd-Giani et al., 2009). To tackle these issues, we have developed the “exploratory Learning Analytics Toolkit” (eLAT). The main aim of eLAT is to support reflection on and improvement of online teaching methods based on personal interests and observations. To help teachers reflect on their teaching according to their own interests, the desired Learning Analytics tool is required to provide a clear, simple, easily interpretable and usable interface while, at the same time, being powerful and flexible enough for data and information exploration purposes. Therefore, eLAT was designed to enable teachers to explore and correlate content usage, user properties, user behavior, as well as assessment results based on individually selected graphical indicators. Dyckhoff et al. (2011) gave a short overview of the toolkit. In the remainder of this paper, we present the theoretical background, design, implementation, and evaluation results of eLAT in more detail.

In section 2 (theoretical background), we provide information on the theoretical background and briefly describe the results of a requirements analysis and its implications for Learning Analytics tools. In section 3 (eLAT: exploratory Learning Analytics Toolkit), we discuss the design, implementation, and evaluation of eLAT. In section 4 (Related Work), we compare our approach to state-of-the-art solutions. Even though there are some approaches to support teachers in their ongoing evaluation and improvement activities (e.g., Johnson et al., 2011a; García-Saiz & Zorilla PantaLeón, 2011; Mazza & Dimitrova, 2007; Pedraza-Perez et al., 2011), many challenges remain (Chatti et al., 2012). Examples include integration with other VLEs and integration of diverse data sources, minimizing the time delay between the capture and use of data, consideration of data privacy issues, protection of students’ identities and prevention of data misuse, enabling data exploration and visualization manipulation based on individual data analysis interests, providing the right information to the right people right away, and investigating which captured variables may be pedagogically meaningful (Elias, 2011; Johnson et al, 2011b). By developing eLAT, we have tried to tackle these challenges. The final section 5 (Conclusion and Outlook) gives a summary of the main findings of this study according to the challenges mentioned above and outlines perspectives for future work.

Theoretical background

In TEL, masses of data can be collected from different kinds of student actions, such as solving assignments, taking exams, online social interaction, participating in discussion forums, and extracurricular activities. This data can be used for Learning Analytics to extract valuable information, which might be helpful for teachers to reflect on their instructional design and management of their courses.

Usable EDM and Learning Analytics tools for teachers that support cyclical research activities are still missing in most current VLE or are far from satisfactory (Hijon & Carlos, 2006). Romero et al. state “[…] data mining tools are normally designed more for power and flexibility than for simplicity. Most of the current data mining tools are too complex for educators to use and their features go well beyond the scope of what an educator might require” (Romero et al., 2007, p. 369). If tracking data is provided in a VLE, it is often incomprehensible, poorly organized, and difficult to follow, because of its tabular format. As a result, only skilled and technically savvy users can utilize it (Mazza & Dimitrova, 2007). But even for them it might be too time consuming. Moreover, unnecessary personal information of students can be observed by teachers or even fellow students, i.e., data privacy issues are ignored in the design of most VLE (Loser & Herrmann, 2009). Legal issues have to be taken into account to prevent malpractice. In Germany, for example, teachers are not allowed to have access to everything a student does in their online courses. They are only supposed to have access to data that is relevant for teaching in a form that is transparent to students (Directive 95/46/EC, 1995; Federal Data Protection Act, 1990). However, many research questions of teachers are concerned with general learning processes of a whole group of students in contrast to gaining more knowledge about a single student. Therefore, to ensure data privacy, student data could be pseudonymized in a preprocessing step by using, e.g., a hash instead of a student-ID, and by presenting summarized results in form of visualizations that rather show group processes and do not allow to focus on one student. Further
deficiencies of reporting tools are related to usability and clarity as well as completeness of the delivered results, such as the lack of possibilities to integrate results of online questionnaires with data from logs.

Many teachers are motivated to evaluate their courses and they already have research questions related to their teaching in mind. For example, a teacher who offers weekly online exercises has the intention to help her students to prepare for an exam. But she is not sure if the currently available exercises are helpful enough for this purpose. Therefore, the teachers would like to know if those students who practice with her online exercises on a continually basis are better in the final exam than students who do not use them. A Learning Analytics toolkit could help her to do research on this hypothesis by automatically collecting, analyzing, and visualizing the right data in an appropriate way. Yet, most monitoring and reporting tools found in current VLEs are designed to collect, analyze, and visualize data in a static tabular form that was predefined by system developers. Teachers face the difficulty that appropriate and usable Learning Analytics tools that help them answer their individual questions continuously and efficiently are missing in prevalent VLEs, since most of the work in the area of Learning Analytics is conceptual (Johnson et al., 2011b).

Teachers should have access to Learning Analytics tools (e.g., provided via dashboards) that can be integrated into a VLE or other learning environments. These tools should allow for interactive configuration in such a way that its users could easily analyze and interpret available data based on individual interests. Results of Learning Analytics and EDM should be presented in a clear and understandable format, e.g., information visualizations that are understandable without data mining expert knowledge. Card et al. (1999) define the term visualization as “the use of computer-supported, interactive, visual representations of data to amplify cognition.” It “promises to help us speed our understanding and action in a world of increasing information volumes” (Card, 2003, p. 542). It has been widely argued in the EDM literature that teachers can grasp the information more easily and quickly when presented through comprehensible information visualizations. Mazza and Dimitrova (2007, p. 138), for instance, write: “…the effectiveness of [Course Management Systems] can be improved by integrating [Information Visualization] techniques to generate appropriate graphical representations ….” Also, results of a recent study by Ali et al. (2012) showed that “visualization can be an effective mean to deal with larger amounts of data in order to sustain the cognitive load of educators at an acceptable level” (p. 486) and “multiple ways of visualizing data increase the perceived value of different feedback types” (p. 488). Still, it should be noted that in some analytical cases visualizations could be ineffective with respect to a textual or a tabular interface, e.g., if details about many items are very important.

In our approach, we only indicate certain facts about the usage and properties of the learning environment and try to visualize them appropriately. Therefore, we revert to the concept of indicators, which can be described as specific calculators with corresponding visualizations, tied to a specific question. For example, if the teacher’s question is “Are those students who practice with online exercise on a continually basis are better in the final exam than students who do not use them,” the corresponding indicator could show a chart that quickly facilitates a visual data comparison. Indicator concepts have been used before. Glahn (2009) for example, introduced the concept of smart indicators, which he defined as “a context aware indicator system, which dynamically aligns data sources, data aggregation, and data presentation to the current context of a learner” (Glahn, 2009, p. 19). However, in our case the target group differs. The eLAT indicators are collecting and visualizing data of students to present them to teachers.

A typical Learning Analytics process is depicted in figure 1. The process starts with the data-gathering step. In this step, data is collected from different learners’ activities when they interact with learning elements within a VLE, LMS or a personal learning environment (PLE). Examples of these activities include participation in collaborative exercises, writing a forum post or reading a document. In the data collection step, it is crucial to address data privacy issues. Often the output of the data extraction and preprocessing step is transferred into a separate database. The second step of the Learning Analytics process is the mining of the preprocessed data, based on different mining techniques, such as clustering, classification, association rule mining, and social network analysis. Thereafter, the results of the mining process can be presented as a widget, which might be integrated into a VLE, a dashboard, or a PLE. Based on appropriate graphical visualizations of the analyzed data, teachers are supposed to be able to more quickly interpret the visualized information, reflect on the impact of their teaching method on the learning behavior and performance of their students, and draw first conclusions about the effectiveness of their teaching, i.e., consider if their goals have been reached. Furthermore, unexpected findings should motivate them to iteratively improve their teaching interventions. However, having a graphical visualization does not guarantee that teachers will be able to
interpret the information represented correctly. Indicators must be designed and evaluated carefully. Also, the system should provide instructions for interpretation.

Figure 1. The Learning Analytics Process

Requirements for developing such dedicated systems have been collected in a former study by analyzing interests and needs of the target group in more detail (Dyckhoff, 2010). Results of this study showed that teachers already have various questions about their instructional design and the utilization of learning materials, the students’ learning behaviors and correlations between objects of teaching and learning as well as outcomes. Their intentions can be e.g., to find out how well the overall instructional design is appreciated, to learn more about the needs of all or a specific group of students, or to better understand learning processes in general.

The conclusion from the study mentioned above was that Learning Analytics tools should support teachers by collecting, integrating, and analyzing data of different sources as well as by providing a step-by-step guidance including semi-automated processes, instead of just presenting large tables of data. “It is undisputable that statistics in isolation represent only one aspect of any real-world situation. To make more meaningful interpretations, [educators] often need to look at the two or more statistics together” (Ali et al., 2012, p. 484). Hence, teachers should be able to choose from a flexible and extendable set of indicators. The system should guide the user throughout the research process, help him or her form research questions, recommend and provide appropriate methods for data collection, integrate data from different sources, and support its collaboratively organized analysis. Such a Learning Analytics tool could e.g., provide extendable lists of supported research questions (indicators) and suitable qualitative as well as quantitative methods for data collection, visualization and analysis. Furthermore, it should be possible to use and integrate the tool with any kind of VLE and learning software.

eLAT: exploratory Learning Analytics Toolkit

In the following sections, we introduce eLAT by giving an overview about results of the requirement analysis, development stages, design decisions, and evaluation phases, concluding with a discussion of our basic findings.

Requirements

Requirements for eLAT have been collected through literature analysis (Dyckhoff, 2010), as well as by informally talking to teachers, eLearning experts and system administrators at RWTH Aachen. The requirements analysis concluded the following main design goals:
• **Usability**: prepare an understandable user interface (UI), appropriate methods for data visualization, and guide the user through the analytics process.

• **Usefulness**: provide relevant, meaningful indicators that help teachers to gain insight in the learning behavior of their students and support them in reflecting on their teaching.

• **Interoperability**: ensure compatibility for any kind of VLE by allowing for integration of different data sources.

• **Extensibility**: allow for incremental extension of analytics functionality after the system has been deployed without rewriting code.

• **Reusability**: target for a building-block approach to make sure that re-using simpler ones can implement more complex functions.

• **Real-time operation**: make sure that the toolkit can return answers within microseconds to allow for an exploratory user experience.

• **Data Privacy**: preserve confidential user information and protect the identities of the users at all times.

**Usability and usefulness**: Every course is different, depending on the teachers and students who are involved in it. There are different teaching strategies, different learning goals, etc. Among the teachers there are some who have not used a Learning Analytics tool before, as well as advanced users. A Learning Analytics tool should be easy to use and understandable for all users. It must be usable for both: the beginner, who just looks at it for the first time, as well as for the expert, who already has a specific question and wants to perform deeper analysis. For beginners, a Learning Analytics tool should enable a direct entry and it should motivate to occupy themselves more with the underlying data, e.g., through a dashboards solution. Experts should find ways to explore and do further analysis to keep them well on the ball. Varying learning scenarios will also demand differing sets of indicators. An important future research task is to find out which indicators are useful for whom in what situation.

**Interoperability, extensibility, and reusability**: Most existing Learning Analytics tools cannot be easily adapted for a different VLE. In addition, new e-learning systems are being developed that may contain useful data for Learning Analytics. Also, learning may take place on informal learning platforms. Therefore, an interoperable Learning Analytics tool that integrates with other systems, and can collect and analyze data from different platforms is required.

**Real-time operation**: New issues on a course may arise at any time during a semester and should then usually be answered directly, so that timely improvements can be made. Also, new questions that are worth to be examined more closely, may arise during the answering process of ongoing questions. Therefore, a Learning Analytics tool should provide current data and comprehensive data analysis capabilities and be available at all times, not only at the end of the semester. Also, interactive analysis and visualization features, like filtering options for exploring the data in more detail, should deliver results and changing visualizations within microseconds.

**Data privacy**: Personal data should be protected at all times to prevent abuse. Data privacy acts ensure such protection (e.g., Directive 95/46/EC, 1995; Federal Data Protection Act, 1990). However, an exception is made for teaching and research projects, under the condition that the data will be handled transparently and purposefully (Federal Data Protection Act, 1990). In addition, students or a data protection officer could be asked to consent to the collection and analysis of student data. Many questions regarding teaching, however, do not aim to examine records of individual students. Rather, data of the totality of students or subgroups with specific characteristics are interesting for drawing conclusions on learning processes. Data could be stored and processed pseudonomized to protect the users. As further protection, the tool could ensure that certain kinds of analyses cannot be executed in certain situations, where they would lead to the identification of individual students.

**Development stages and evaluation methods**

eLAT was iteratively and incremental developed within two main stages that partially overlapped to meet the requirements described above: (stage 1) the implementation and testing of a backend framework as well as (stage 2) the design and evaluation of a UI (frontend).

In the first stage eLAT was designed as a prototype to evaluate different software architectural approaches for Learning Analytics using different VLE platforms. During winter term 2010/2011, we selected four courses that were using the learning and teaching portal L²P of RWTH Aachen University. The courses differed in course sizes
(1370, 338, 220, and 38 registered students), learning technologies and teaching styles to ensure realistic usage scenarios. We logged the students’ activity, interaction and (in one case the) assessment data over the duration of three months. By using the data of real courses, it was possible to learn more about meaningfulness of already implemented indicators and to let the teachers of the courses participate in the development process of eLAT. In this way, we could get immediate feedback and comments on prototype stages that already processed analytics based on the real data.

The design and evaluation of a UI (stage 2) started parallel to the first development stage described above. It was iteratively conducted as well, whereat each of overall three iterations had a specific objective. The first iteration dealt with the collection and definition of the content. Since eLAT was designed to enable teachers to explore educational data of their students and courses based on graphical indicators, it involved the collection of indicators as well as assigning priorities to them. Thus, semi-structured interviews were conducted to evaluate a set of graphical indicators and to get to know further user requirements. Semi-structured interviews are used to collect facts, opinion and attitudes of the interview partners (Naderer, 2007). The interviews are guided through prepared questions, but it is also possible to ask questions spontaneously to investigate interesting details (Lindlof and Tylor, 2002). The second iteration focused on the layout and data presentation of the UI. The evaluations of the first and second iteration were performed with the help of paper prototypes, while in the third iteration a functional UI, which was implemented based on previous evaluation results, was used to investigate interactivity and usability aspects.

Layout and data presentation were designed and evaluated in terms of heuristic evaluation, cognitive walkthrough and pluralistic walkthrough. A heuristic evaluation uses approved usability principles or guidelines to investigate the usability of a UI. Thus, problems can be discovered with less effort in an early development step (Nielsen, 1992). A cognitive walkthrough is more formal than heuristic evaluation. It needs a specification of the UI and tasks to evaluate the usability. With the help of the tasks the UI can be processed step by step to discover usability problems (Polson et al., 1992; Dix et al., 2004). Both methods were chosen to evaluate the prototypes of the UI in an early stage of development. The pluralistic walkthrough is similar to the cognitive walkthrough. It is a meeting of experts of different domains, such as users, designers, and usability specialist. They discuss elements of the interface prototype according to the view of the users (Bias, 1994). We used a variant of the pluralistic walkthrough where a domain expert, a usability expert and the designer discussed the interface from the users’ perspective. Main results of these studies are presented in the section “User interface”.

The third iteration, which was mainly concerned with interaction, included a qualitative think-aloud study. Here users were asked to perform tasks with a software prototype, whereat they were talking about what they were doing and thinking. During the tasks the evaluator observed them. This method was chosen to identify areas of interactions where users can make mistakes (Dix et al., 2004).

In the following sections, we present the resulting UI, use cases, design and implementation details of eLAT as well as overall evaluations results.

User interface

The structure and layout of the eLAT user interface (UI) are the result of an iterative approach and were derived from our user studies, which have been elaborately discussed in Bültmann (2011). The UI is designed as a launch pad, which is similar to a dashboard but provides more comprehensive analysis options, additional to an initial overview (Few, 2006).

A monitoring view helps to observe several indicators at once (figure 2). Furthermore, analysis views provide a deeper insight into the data of chosen indicators by making it possible to drill down into details by changing parameters of an indicator. Additionally, a mouse over effect shows details about the currently regarded information (figure 3).

In the monitoring view, the content of the launch pad is grouped into four widgets. The widgets are containers for indicators related to the categories “document usage,” “assessment/ performance,” “user activity” and “communication.” Each indicator has its own tab in the widget. This hierarchical layout is supposed to help users to get a better overview about the current learning situation. By using widgets and tabs it is possible to put all indicators...
on one single screen. This concept also helps in terms of personalization, because widgets can be arranged flexible by the users.

Figure 2. Monitoring view of the eLAT user interface

Figure 3. Analysis view of the indicator “Activity behavior”

The analysis view of each indicator is consistently accessible in the corresponding tab by clicking “Configure indicator details” (figure 2). This detailed view of the indicator is shown as an overlay on top of the monitoring view (figure 3). Layout and functionality have been designed in a consistent way to gain a better usability (Few, 2006). On the right side of the analysis view is a filtering menu. The filtering of the presented data is context-dependent according to the currently selected indicator. For each tab in the filtering menu of any indicator, the user can determine which information the indicator should present. Hence, there are many options for data exploration, such as, comparing the activity of male and female users or the activity of students of different study programs. But not all filters can be used for each context. Because of data privacy regulations, we cannot allow the use of any kind of user properties like gender or study course, when there are less than a certain number of students, e.g. five users with that...
certain property in a course. The following paragraphs give an overview about six implemented indicators, which were rated to be interesting as result of the evaluations.

Figure 3 shows the analysis view of the indicator “Activity behavior”. Student data is divided into the three groups “very active students” (blue bars), “active students” (red bars), and “inactive students” (yellow bars), and shows their weekly distribution over a time span. An “active student” is determined by calculating the average number of days per week, at which a student was active, i.e., logged into the system. In the current configuration of the indicator “activity behavior”, shown in figure 3, a student is defined to be “active” if he or she logs in at least once a week. A student is defined to be “very active” if he or she logs in on more than five days a week. The user can change the time span and the definition of an “active student.”

The data in figures 3–7 is based on a programming course, which was finished with a final exam at February 8th 2010. As expected, the indicator above shows an increase of “inactive students” after the exam date. The indicator “Activity behavior” (figure 3) indicates whether continues learning is taking place. A participant of our semi-structured interviews considered continues learning as a main factor for good exam results. As a sign of continues learning, the teacher might e.g., expect his students to log in at least twice a week to download new materials and stay up-to-date related to course information. The indicator “Activity behavior” can show tendencies of increasing or decreasing numbers of such active (groups of) students. High numbers of inactive students during the semester could bring the teacher to motivate his students to learn more regularly, e.g. through creating weekly exercises, or initialize further investigations on the reasons of low activity.

The indicator “access and accessing students” (figure 4) supports the teachers in monitoring the overall online activity of their course. It shows the number of accesses/clicks (blue line) over the number of unique students (red line) who accessed the virtual learning environment during a time span defined by the user. The blue line represents the sum of every single click on any resources in the learning environment per day or week. It is important for a teacher to observe if e.g., a small group of students clicks many times or many students click once on a resource. The data in figure 4 shows that almost every day about a third of 278 registered students accessed several resources. The peak at the end of the timeline demonstrates a strong increase in accesses before the final exam, but only a small increase in accessing students. Lines converge after the date of the exam. Probably, the students only come back to the virtual course room to check the exam results (one click per student). The indicator “Access and accessing students (figure 4) can show outliers of usual access behavior/frequency. Teachers can relate high or low usage e.g., to teaching events or holidays. They can quickly observe if changes of learning materials or didactics lead to changes in overall usage behaviors. This might motivate them to experiment with didactics to improve the overall access to the learning environment.
With help of the “Activity areas” indicator, shown in figure 5, teachers are supposed to identify whether and when students are accessing which parts/areas of a virtual course room per week in a defined time span. Hence, access rates between functions, like wiki pages or discussion forum, can be compared and related to teaching events as well. The x-axis of the indicator shows the days or weeks. If metadata on dates of course events, like the occurrence of specific lectures or the exam, are provided, these events can also be written on the x-axis. The y-axis records the number of students, who were active, i.e., clicked on resources, during that day or week in a specific part of the virtual course room. The red line in figure 5 e.g., shows the number of students accessing a document library with learning materials, such as lecture scripts and exercises. The red line and the yellow line, which represents the number of students, who accessed the discussion forum, peak out 1–3 day before the exam. Students seem to become more active in reading and discussing during that time, so that the case could be made that they are learning more intensively.

The “Top 10 resources” indicator (figure 6) gives an overview about the most accessed materials. It can help to identify active documents/items that have been accessed more than others. Such a popularity indicator could have differing reasons. A document that shows up in the “Top 10 resources” indicator e.g., could be useful for solving an exercise or might be difficult to understand. The learning materials, presented in figure 6, are exercises (“Uebung 10–13”), code examples (“10-class”), lecture scripts (“12-GUI” and “11-Exc”), a lecture summary (“14-wdhlg”) and an example solution (“Loesung 10”). This could indicate that students mainly have been learning by solving exercises. Based on this indicator, a teacher could start to explore the meaning of the high access of specific learning materials. In case of a difficulty of understanding, learning materials could be improved.
The “Forum usage” indicator (figure 7) represents the number of new threads with corresponding answers to these threads (x-axis) per day (y-axis). Teachers can more easily identify increasing discussions and, thus, determine, if collaboration among students is taking place, and whether there might be problems or not. Furthermore, by looking at a thread title of the observed communication activity, problems in understanding or in preparation of learning materials could be identified. Although this indicator does not show the answers per thread it can be an activity measure for forum usage.

The “Adoption rate” indicator (figure 8) deals with the time span from uploading a selected learning material to the time of access by students. With the help of the “adoption rate”, it can be identified how fast how many students access new materials. It also shows the number of students who have utilized a certain material and helps teachers to find out after which time the item has achieved a sufficient distribution amongst his or her students. In some courses this is helpful e.g., because the students might have a reading assignment. Thus, teachers are enabled to estimate,
how many students have at least accessed a document that they were supposed to read. If the adoption rate is lower than expected, this could be an explanation for low homework discussion participation during class.

Comprehensive and well-fitting indicators are crucial for an effective and successful application of a Learning Analytics tool. Therefore, we conducted semi-structured interviews with eight teachers (6 male, 2 female). During the interviews, example indicators were presented, shortly explained and discussed with the teachers to collect facts, opinion and new ideas. The evaluation goals named by all participants were quality and activity oriented. It is important for seven interview participants to monitor activity, like the frequency of logins or the date and occurrences of access. The principal point is to monitor, if continuous learning is taking place and if this leads to better learning results (mentioned by all participants). Therefore, it is important to relate the usage and access of learning material with performance. These evaluation goals are similar to the results found in our former study (Dyckhoff, 2010). Indicators concerning activity or achievements are valued most.

Seven participants rated the indicator “Access and accessing students” (figure 4) as helpful for a general overview, especially in the beginning of using a VLE. But the posed question, if it can always deliver interesting information, remained unanswered. Therefore, the indicator should be used in combination with other data. Six participants stated that “Adoption rate” (figure 8) is important. It can indicate student behavior. Teachers can relate it to teaching events, such as making an announcement, and observe changes in student behaviors. The “Top 10 resources” indicator (figure 6) was identified as a valuable measure (7 called it important and one somehow important). It helps to identify those resources that are somehow relevant for students. A “Top 20” might be better suited for courses with a large amount of resources. By monitoring, e.g., “Activity areas” (figure 5), it is possible to identify patterns, such as at what times students do their exercises, or whether they access learning materials before or after lectures. Activity measures in general were pointed out to be well suited for first impressions during analysis, i.e., for getting a quick overview about what is going on.

Regarding the performance, besides correlations with the activity of students, teachers also wished to take a deeper look at correlations with properties, like the program of study, the duration of study, and the mother languages of students. Five of eight participants stated that such indicators are important and one thought it is not important. This indicator could be important for adjusting teaching methods for specific groups of students.

Teachers had divergent opinions on examining active participation, e.g. the usage of forums or wiki pages (figure 7). Four participants rated it as an important measure, because communication, discussion and participation are represented in collaborative features. The other half rated it less important, because of a low participation rate in their courses or students using other collaboration tools in the cloud. The assessed usefulness of this indicator heavily depends on the participation of the students in an online course and the relevance a teacher ascribes to it. The reason can be found in the underlying structure of the hybrid courses. Communication often takes place outside the VLE because of blended learning settings at RWTH Aachen University. Students still talk and learn together outside the online learning environment, and thus, communication and its relation to learning cannot be measured adequately.

Additional to the indicators described above, we evaluated several other indicators. The final prototype does not implement all indicators evaluated in the previously conducted iterations due to the fact that not all the data needed was available at the time of implementation; e.g., it lacks of data corresponding to session information and durations or more detailed metadata about students. Those indicators, not described in this paper, have let to differing opinions. Their usefulness rating was low or very much dependent on the underlying didactical scenarios. Hence, we draw the overall conclusion that teachers should be enabled to explore data individually, e.g. by arranging their own sets of indicators per course, to facilitate improvement of teaching.

Framework design and implementation

The architecture of eLAT is presented in an abstract manner in figure 9. The functional requirements for the toolkit demanded a very flexible coupling of VLE infrastructure, the implementation of an indicator evaluation process and a visualization system. eLAT encompasses three main components, namely an indicator framework, a mining database, and a visualizer application. At the heart of eLAT lies the indicator framework, which negotiates and provides report evaluation services, i.e., executions of indicator calculations, between the website and the mining
database, while the visualizer component provides an abstraction layer for visualizing different reports in an appropriate manner.

![Diagram of eLAT system architecture]

The implemented course of action for an indicator evaluation is illustrated in the green boxes in figure 9. Triggered by the instructor who is visiting the website and selecting an appropriate indicator (step 1), a Controller Factory in the indicator framework will then dynamically instantiate a controller for this indicator. This will create a view, containing user interfaces for all the available configuration properties (step 2). After the user has finished adjusting all the properties and the configuration has been validated, the framework will generate a report evaluation request, store it in the report database, and send an evaluation request to the evaluation service instances. In the meantime the user will be redirected to a waiting page that displays the current evaluation status and updates automatically (step 3+4). Once the evaluation has completed the report, consisting of the initial configuration and a dataset of raw data tables, it will be stored permanently in the report database (step 5). While querying the report status the client side scripts will eventually learn of the successful evaluation and load the dedicated JavaScripts that will then generate and show the appropriate visualization based on the raw data set obtained by the report service (step 6+7).

A very important part of the system architecture of eLAT deals with the extensibility and reusability of the existing code base, so that the scope of operations can grow. Therefore, a single indicator implementation makes use of smaller parts in the form of expressions that are performance-optimized database queries to retrieve specific result sets that can be useful for other indicators as well. The same practice is applied to the dynamic user interface generation for indicator parameters and the visualizers, which operate on standardized datasets and are therefore generic to the data inside the report. This leads to a relatively small effort for implementing new indicators.

Since we want to keep the eLAT implementation independent from a particular VLE, we have developed a neutral data model that supports all the major data types as well as an extension model to fit in special types. As illustrated in figure 9, we have modeled a learning environment as a general virtual learning space dedicated to a specific course, which could be implemented in any VLE. From the VLE instance we extract the information needed for further investigation and load them into the mining database:

- **User data**: This primarily stores the pseudonymous user identifier and the role of the user, e.g. “instructor”, “tutor” or “student”. It also holds custom properties like gender and study course, which might not be available for all users.
- **Content data**: In most VLE, there are multiple sources of content data, structured by using content areas and content lists or libraries, which can have predefined recurring types and certain properties for specific objectives. Additionally, there are content type extensions dedicated to extend content items with properties that are not
always available, like for example a forum post, which has a property for the depth inside the discussion thread – in any other case that post can be regarded like any other content item, with properties like title, creation date and the user responsible.

- **Assessment data:** We wanted to store assessment instances separately from the user submission to allow for different handling and support of various assessment types. This also supports an extension model to fit in properties like completion time or group work submissions, which are not common, but sometimes available.
- **Activity data:** This contains specific information about any content interaction along with timestamp and user information occurring in the learning environment.
- **Event data:** As typically every course has a calendar and special events like course start, assessment and exam dates, it is useful to use this information in the parameter selection and visualization process. We are interested in narrowing down the activity to certain time spans such as the time between two exams or visualizing the usage statistics of the forum activities only in the time between two evaluations of weekly assignments.

Most of the server-side code is written in .NET and uses Windows Communication Foundation (WCF) services for providing data and communication interfaces between the website, the evaluation service and the client site visualizations. The indicator website itself does not generate much load on the CPU or the database (answers below half a second), relative to the load the evaluation processes. This is why we decided to execute evaluation processes by a dedicated service that is running in multiple instances, ideally on a different physical machine, to approach our real-time operation requirements.

As a result of the semi-structured interviews, the possibility to filter data was identified as a substantial requirement. However, when implementing the filtering options, we did not yet entirely meet our goal of providing real-time operations, which are necessary for a smooth user experience. One reason for this lies in the structure of the database, which was created in the first development stage and, in its current form, has been designed according to a pattern used for operational databases rather than data warehouses. Hence, we implemented on-the-fly generation of lookup tables, containing pre-calculated intermediate results, to increase the efficiency of filtering. However, a thorough redesign of the database scheme with data warehouse principles in mind is required to address this issue in greater detail.

To allow for upgrading and enhancing of analysis methodology while the system is deployed, the implementation makes heavy use of the Managed Extensibility Framework, which allows for dynamic composition and object instantiation of software capabilities and is part of the latest version of the .NET Framework.

For our software prototype evaluations we extracted data for our four courses from L²P, which is currently run on SharePoint 2007, and one a Moodle instance, which was used in one course to host assessments. We employed custom code for each VLE to extract and transform data to our data model, as shown in the blue boxes in figure 8. From there the simplest way to load data into the mining database is by using the xml import feature. We have specified an xml schema, which can be used to create and validate xml files from any VLE data output.

To ensure data privacy, we had to make modifications to the data extraction process. Since we have no interest in the identity of each user, but still need to be able to uniquely distinguish their data, we chose to create a hash from the username with a learning environment specific salt, so that the transformation is not easily reversible. The algorithm used is MD5 with a hash size of 128 bits (Rivest, 1992).

In future versions, we still expect the need for custom data exports of indicator reports. Some advanced users have already expressed their interest in using the results in Excel or other analytics tools, like Weka (Hall et al., 2009) or Keel (Alcalá-Fdez et al., 2009), to start their own studies. Although we would like to provide as much functionality as possible inside the eLAT toolkit, we have designed interfaces for exporting reports.

**Evaluation and discussion**

This section presents our basic evaluation findings and answers the two questions: Does eLAT meet the requirements “usability,” “usefulness,” “interoperability,” “extensibility,” “reusability,” “real-time operation” and “data privacy” sufficiently and what did we learn in relation to those challenges.
To evaluate if all requirements are met by our concept of a usable exploratory Learning Analytics toolkit, we have implemented an instantiation of our concept, namely the prototype eLAT, tested its technical functionality with four courses at RWTH Aachen University and conducted several usability evaluation iterations. Results of the first two iterations have been summarized in the section “User interface” by describing the UI and giving insights into indicator usefulness ratings.

The final usability study with the software prototype was conducted as a qualitative think aloud study. Four teachers (3 male, 1 female), participants of the former iterations, were asked to perform tasks, which aimed at keeping an eye on the learning progress. At first, the prototype was shown to the participants without giving a detailed explanation. The tasks were designed with increasing difficulty. All participants quickly understood the structure and were able to navigate the user interface. The monitoring and analysis views were well understood. One out of four participants tried to click and investigate the whole interface. During the evaluation, some problems were discovered.

One of the participants criticized the font size as being too small, but put into perspective that the font in the web browser could be enlarged by him and thus individually adopted. Furthermore, the mouse over effect, which shows details on the currently regarded information, helped to overcome the drawback of a small font. Two participants did not immediately understand the wording of the indicator “Adoption rate” (figure 8). Regarding the filtering, three participants requested more analysis options than possible. One wished the possibility to get into details by simply clicking on a line of a chart. As an improvement for filtering options, one participant mentioned the possibility of providing templates for common time spans during the lecture period, such as “examination phase” or “examination preparation phase”. Two participants stated that the tool should be extended regarding the correlation of activity and performance. Also the filtering possibilities should be extended and some filters also could be applied in other indicator contexts. Furthermore, filters according to the activity behavior of students could be defined. In addition, advanced users still need more personalization and more analysis functionality, e.g. the possibility to create new indicator templates.

However, although the evaluation showed that eLAT is usable, comprehensive field-tests with more courses from different disciplines still have to be carried out to gain more reliable data on pedagogical usefulness of the toolkit. We investigated the usability of several indicators. Yet, we do not have enough data to answer the question which captured variables may be pedagogically meaningful sufficiently. Indicators implemented in eLAT are rather general and focused on simplicity and understandability. For enabling data exploration and visualization manipulation based on individual data analysis interests, more significant indicators that are still easy to use, have to be designed, implemented, and evaluated. Reviewers suggested e.g., considering outliers of student behavior by using a box plot for visualization. May et al. (2010) suggested potentially meaningful indicators, such as, indicators for visualizing different user activities as unique colored spheres on a horizontal bar, or a radar graph for visualizing different aspects of the user’s level of interaction in a communication activity. Zorilla et al. (2010) chose a radar graph as well for presenting student and session profiles. Yet, the authors of both research projects mention evaluated difficulties in the understandability caused by the visualizations (May et al., 2010; Zorilla et al., 2010). These evaluation results conform to findings from our indicator evaluation: visualization types should be well known and simple to interpret. Fancy graphics might cause problems by requiring too much interpretation time, but they have the potential to deliver meaningful information. The most impact has a visualization that may present outliers of data or rather unexpected data about students. Therefore, the design of indicators should be carefully accompanied by user studies and supplemented by adequate interaction options and help facilities as needed.

For providing the right information to the right people, the development of more advanced functionality, like the systematic comparison of selected data of different courses should be pushed on. As a further advancement, the tool could analyze indicator usage and offer a sophisticated rating mechanism to recommend indicators depending on teachers’ data analysis goals. In addition, eLAT could expand its target group to learners for the purpose of self-monitoring.

Current Learning Analytics tools should be interoperable with different learning environments and systems. eLAT uses a neutral data model and has been tested with data of three different learning environments, namely the learning and teaching portal L²P, Moodle and Dynexite, an e-learning exercise tool. We are also working on a mapping between our data model and Contextualized Attention Metadata (CAM) (Schmitz et al., 2009) that will be supported in the next versions of eLAT. In the future, Learning Analytics tools may be linked more often to more open, networked, and lifelong learning environments, driven by new learning theories, such as Connectivism (Siemens,
Hence, integratability is an important factor for the sustainability of Learning Analytics tools. Furthermore, integration of diverse data sources may lead to more pedagogically meaningful indicators because a more holistic picture of the learner could be drawn. To support the implementation of new indicators, in terms of extensibility and reusability, indicator implementations make use of expressions that are performance-optimized database queries. Furthermore, visualizers operate on standardized datasets and are therefore generic to the data inside the report.

As a result of the semi-structured interview, the possibility to filter data was identified as a substantial requirement. However, when implementing the filtering options, we did not entirely meet our goal of providing real-time operations, which are necessary for a smooth user experience. The reason can be found in the data tables, which are not normalized in order to improve performance. To avoid too long calculations, we implemented on-the-fly generation of lookup tables containing pre-calculated intermediate results, thus, increasing the efficiency of filtering. However, a thorough redesign of the database scheme with data warehouse principles in mind is required to minimize the time delay between the capture and use of data.

Regarding data privacy, we chose to pseudonymize student data to protect students’ identities and prevent data misuse. According to prior agreements with the universities data protection officer, creating a hash from the username pseudonymized the collected data for clearly assigning and saving personal data. We used a learning environment specific salt with the username, which is a value that makes it difficult to guess the original value, for security reasons. Data was exported from the VLE to eL²P’s backend on a weekly basis during the semester. For transparency reasons, students of each course were informed verbally and in writing about the data collection and analysis goals. Of course, the more data is collected, the higher is the potential to recognize individual students during the data analysis. Therefore, we prohibit certain parameter selection options, if less than five students with that particular property are registered for a course. This leads to a trade-off of “data privacy” versus “pedagogical useful indicators”. Teachers of small courses with few registered students will not have as much indicator options as large courses with many students. Also, courses with unequally distributed student properties, e.g. with very few female students, cannot be analyzed according to that property, i.e. questions on gender differences cannot be investigated.

**Related Work**

Several researchers are trying to solve some of the EDM and Learning Analytics challenges similar to the ones mentioned above.

Krüger et al. (2010) observed that VLEs are usually not designed for data analysis and mining. Therefore they also developed a data model to allow for efficient data mining in a VLE. A prior aim was to automate and alleviate the pre-processing, which is needed to explore, analyze, and mine VLE data. The data model and an implementation are oriented towards the structure of Moodle and have not yet been tested with other VLEs. Our own data model design differs from this work in terms of more detailed logging data concerning content interactions, a property bag mechanism for storing custom data type extensions, the introduction of events as its own data type and an assessment data model. Reasons for these differences are due to our requirements that demand the possibility for integrating custom content and user properties that diverge with different course type instances as well as the need for adding versatile information. An extension model enables instructors to extend the information set of his or her students by using surveys inside the course room, while preserving data privacy, and using the results and user submissions for cross-comparison in certain indicators. Furthermore, our data model was designed to fit with different kinds of VLEs and has been evaluated with L²P as well as Moodle.

Pedraza-Perez et al. (2011) have presented a first prototype of a java desktop tool addressed to non-data-mining experts, which supports simple execution of common data mining steps. A simple wizard-based interface helps the user to create, pre-process, visualize, and mine Moodle data files by clicking buttons that are named accordingly, e.g. “create data file” or “pre-process data file”, whereby they choose between pre-defined and recommended data mining options. Even though its usage is simple, the interface is still oriented towards users with background knowledge about the data mining process. Teachers, who have no data mining experience, will first have to learn about the process and understand the wizard’s vocabulary, such as “classification,” “regression,” or “association.”
Lately, several research projects similar to eLAT are emerging. These projects show that information visualizations are gaining importance for increasing usability of Learning Analytics tools. Mazza and Dimitrova (2007) studied graphical user interfaces for EDM tools. They presented CourseVis, which has been built as an extension of WebCT. Its design is based on the results of a survey, which revealed that instructors need information on social, cognitive, and behavioral aspects about their students when running distant education courses with a VLE. CourseVis uses web log data similar to eLAT and also renders it graphically. An evaluation has shown that the graphical representations of CourseVis helped instructors to quickly and more accurately grasp information of students (Mazza, 2006). As a follow-up, the successful visualization principles of CourseVis have been implemented with a graphical interactive plug-in for student monitoring in Moodle, called GISMO (Mazza and Botturi, 2007). Another example is EDM Vis, an information visualization tool for exploring students’ data logs that uses a tree structure to provide an overview of class performance, and interface elements to allow easy navigation and exploration of student behavior (Johnson et al., 2011a). The main differences of CourseVis, GISMO, and EDM Vis compared to eLAT are that the three former mentioned provide visualization in a more static way defined by the system developers. In our approach, teachers are able to dynamically choose by themselves which visual indicators are most helpful to answer their individual data analysis questions. They can change parameters of existing visualization and, hence, create sets of relevant indicators composed on dashboards that fit a specific teaching intervention.

Based on motivations and goals very similar to those of eLAT, two other tools are taking individual user perspectives into account: Graf et al. (2011) recently introduced the Academic Analytics tool (AAT) and García-Saiz and Zorilla Pantaleón (2011) presented the data mining application, called E-learning Web Miner (EIWM). AAT has been primarily developed for learning designers, but can also be used by teachers. The application allows its users to access and analyze students’ behavior in Moodle online courses by customizing and performing analytical data base queries. This data analysis may be used as additional evidence for the formative evaluation of courses. Combined with students’ evaluations and professor and tutor recommendations for changes, the analytics results are supposed to inform the work of learning designers, who can then adapt, revise and extend resources. Like eLAT, AAT is applicable for different VLEs and aims at being easily extendable, with respect to adding analysis techniques/indicators (Graf et al., 2011). EIWM is supposed to help teachers explore distance students’ behavior in a VLE, e.g., Moodle or Blackboard. It offers a set of templates, which can be compared with eLAT’s indicator approach. Currently these templates visualize data graphically related to three common questions of teachers, concerning resource usage, students’ sessions, and students’ profiles in a selected course. The templates contain all needed input attributes for the execution of the related data mining algorithms, which have been implemented on the basis of the Weka toolkit (García-Saiz & Zorilla Pantaleón, 2011). A challenge for Learning Analytics tools like EIWM or eLAT is not only to provide meaningful templates/indicators, but also to decrease the necessity to input parameters by users. As distinguished from EIWM currently eLAT only implements indicators based on simple statistical methods. In the future, we are going to implement further indicators based on data mining methods and evaluate their usefulness for facilitating analysis and improvement of teaching.

It is a complex task for teachers to input data mining parameters. Therefore, parameter-free data mining is a promising approach. Zorilla et al. (2011) compared yacaree, a parameter-free association miner with three well-known association rule miners. In their study yacaree was well-suited and superior in terms of usefulness to a teacher involved in the evaluation. Research like this will support the further development of usable Learning Analytics and EDM tools that support explorative analytics usage.

**Conclusion and outlook**

In this paper, we presented the theoretical background, requirements, design, implementation, and evaluation details of eLAT; an exploratory Learning Analytics Toolkit that enables teachers to monitor and analyze their teaching activities. The main goal of eLAT is the improvement of teacher support with graphical analytics, which are useful because they allow extending the audience to “normal” instructors without prior knowledge in data mining techniques. With the help of eLAT, teachers are enabled to explore, reflect and evaluate teaching interventions based on their interests. Key EDM and Learning Analytics requirements, such as usability, interoperability, extensibility, reusability, and data privacy, have been tackled with the development of eLAT.

Currently, eLAT has been primarily developed with the intention to support teachers in their ongoing reflection, evaluation and improvement of their instructional design. In the future, we plan to enhance eLAT in ways that...
students can use it as well. Furthermore, eLAT has been successfully tested with data collected from four courses. Future work will include the integration of eLAT in L²P and its field-testing with more courses from different disciplines, based on new indicators. We also plan to enhance eLAT with an intelligent recommendation component and evaluate its usefulness.

Acknowledgements

This project was partly funded by the Excellence Initiative of the Federal and State Governments.

References


May, M., George, S., & Prévôt, P. (2010). Travis to enhance online tutoring and learning activities: Real time visualization of students tracking data. In M. B. Nunes & M. McPherson (Eds.), *Proceedings of the IADIS International Conference E-Learning 2010* (pp. 57–64). Freiburg, Germany: IADIS.


Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade

M'hammed Abdous1, Wu He2* and Cherng-Jyh Yen

1Center for Learning and Teaching // 2Information Technology Department of Information Technology & Decision Sciences // 3Educational Research and Statistics Department of Educational Foundations and Leadership, Old Dominion University, Norfolk, Virginia, 23529, USA // mabdous@odu.edu // whe@odu.edu // cyen@odu.edu

*Corresponding Author

ABSTRACT

As higher education diversifies its delivery modes, our ability to use the predictive and analytical power of educational data mining (EDM) to understand students’ learning experiences is a critical step forward. The adoption of EDM by higher education as an analytical and decision making tool is offering new opportunities to exploit the untapped data generated by various student information systems (SIS) and learning management systems (LMS). This paper describes a hybrid approach which uses EDM and regression analysis to analyse live video streaming (LVS) students’ online learning behaviours and their performance in their courses. Students’ participation and login frequency, as well as the number of chat messages and questions that they submit to their instructors, were analysed, along with students’ final grades. Results of the study show a considerable variability in students’ questions and chat messages. Unlike previous studies, this study suggests no correlation between students’ number of questions / chat messages / login times and students’ success. However, our case study reveals that combining EDM with traditional statistical analysis provides a strong and coherent analytical framework capable of enabling a deeper and richer understanding of students’ learning behaviours and experiences.

Keywords

Educational data mining, Data mining, Live video streaming, Clustering analysis

Introduction

According to a recent survey conducted by Campus Computing (campuscomputing.net) and WCET (wcet.info), almost 88% of the surveyed institutions reported having used an LMS (Learning Management System) as a medium for course delivery for both on-campus and online offerings. In addition to various student information management systems (SISs), LMSs are providing the educational community with a goldmine of unexploited data about students’ learning characteristics, behaviours, and patterns. The turning of such raw data into useful information and knowledge will enable institutes of higher education (HEIs) to rethink and improve students’ learning experiences by using the data to streamline their teaching and learning processes, to extract and analyse students’ learning and navigation patterns and behaviours, to analyse threaded discussion and interaction logs, and to provide feedback to students and to faculty about the unfolding of their students’ learning experiences (Hung & Crooks, 2009; Garcia, Romero, Ventura, & de Castro, 2011). To this end, data mining has emerged as a powerful analytical and exploratory tool supported by faster multi-core 64 CPUs with larger memories, and by powerful database reporting tools. Originating in corporate business practices, data mining is multidisciplinary by nature and springs from several different disciplines including computer science, artificial intelligence, statistics, and biometrics. Using various approaches (such as classification, clustering, association rules, and visualization), data mining has been gaining momentum in higher education, which is now using a variety of applications, most notably in enrolment, learning patterns, personalization, and threaded discussion analysis. By discovering hidden relationships, patterns, and interdependencies, and by correlating raw/unstructured institutional data, data mining is beginning to facilitate the decision-making process in higher educational institutions.

This interest in data mining is timely and critical, particularly as universities are diversifying their delivery modes to include more online and mobile learning environments. EDM has the potential to help HEIs understand the dynamics and patterns of a variety of learning environments and to provide insightful data for rethinking and improving students’ learning experiences.

This paper is focused on understanding live video streaming (LVS) students’ learning behaviours, their interactions, and their learning outcomes. More specifically, this study explores how the interaction of students with each other and with their instructors predicts their learning outcomes (as measured by their final grades). By investigating these...
interrelated dimensions, this study aims to enrich the existing body of literature, while augmenting the understanding of effective learning strategies across a variety of new delivery modes.

This paper is divided into four sections. It begins by reviewing the literature dealing with the use of data mining in administrative and academic environments, followed by a short discussion of the way in which data mining is used to understand various dimensions of learning. The second section explains the purpose and the research questions explored in this paper. The third section describes the background of the study and details its methodological approach (sampling, data collection, and analysis). The paper concludes by highlighting key findings, by discussing the study’s limitations, and by proposing several recommendations for distance education administrators and practitioners.

Data mining applications in administrative and academic environments

At the intersection of several disciplines including computer science, statistics, psychometrics (Garcia et al., 2011), data mining has thrived in business practices as a knowledge discovery tool intended to transform raw data into high-level knowledge for decision support (Hen & Lee, 2008). To this end, a wide range of tools that can be used for collecting, storing, analysing, and visualizing data, such as the SPSS Modeler (formerly Clementine) and the SAS Enterprise Miner, have been developed in the business world. These tools use sophisticated computing paradigms including decision tree construction, rule induction, clustering, logic programming, and statistical algorithms.

Although data mining has been widely used in business environments to predict future trends and consumer behaviours (Harding, Shahbaz, Srinivas, & Kusiak, 2006; Ngai, Xiu, & Chau, 2009), the data mining method has been dramatically under-used in education research in general (Faulkner, Davidson, & McPherson, 2010). Only recently have higher education institutions started to exploit the potential of this powerful analytical tool (Black, Dawson, & Priem, 2008).

However, according to Romero and Ventura (2010), educational data mining (EDM) has emerged as a new field of research capable of exploiting the abundant data generated by various systems for use in decision making. The enthusiastic adoption of data mining tools by higher education has the potential to improve some aspects of the quality of education, while it lays the foundation for a more effective understanding of the learning process (Baker & Yacef, 2009). EDM, when integrated into an iterative cycle (Romero, Ventura, & Garcia, 2008) in which mined knowledge is integrated into the loop of the system not only to facilitate and enhance learning as a whole, but also to filter mined knowledge for decision making (Romero et al., 2008) or even to create intelligence upon which students, instructors, or administrators can build, can notably change academic behaviour (Baepler & Murdoch, 2010).

From an administrative perspective, Chang (2006) argues that the predictive capacity of data mining can further enhance enrolment management strategies by increasing the HEIs’ understanding about their admitted applicants. Similarly, Delavaria, Phon-Amnua isuka, and Reza Beikzadeh (2008) contend that data mining knowledge techniques are capable of enabling higher learning institutions to make better decisions, to put more advanced planning into place to direct students, and to predict individual behaviours with higher accuracy, and, in so doing, to enable the institutions to allocate resources and staff more effectively. Without inflating the merits of data mining in rethinking administrative and academic processes, it is clear that data-mining is gaining ground and is providing powerful analytical tools capable of converting untapped LMS and EPR data into critical decision-making tools with the potential of enhancing students’ learning experiences (Garcia et al., 2011).

From a learning perspective, according to Castro, Vellido, Nebot, and Mugica (2007), data mining is being used in higher education

- to assess students’ learning performance
- to provide feedback and adapt learning recommendations based on students’ learning behaviours
- to evaluate learning materials and web-based courses, and
- to detect atypical students’ learning behaviours.

Following this line of thinking, Perera, Kay, Koprinska, Yacef, and Zaiane (2009) used clustered data mining techniques to support the learning of group skills by building automated mirroring tools capable of facilitating group
work. In a similar study, Sun, Cheng, Lin, and Wang (2008) used rules based on data mining results to form high interaction-learning groups.

For their part, Hung and Zhang (2008) applied data mining techniques to server logs, both to reveal online learning behaviour patterns and to support online learning management, facilitation, and design. Their study’s results revealed students’ behavioural patterns and preferences, which helped them to identify active and passive learners and which extracted important parameters for the prediction of the students’ performance (Hung & Zhang). Using a similar approach, Ba-Omar, Petrounias, and Anwar (2007) analysed web access logs to identify learning patterns and offline learning styles. In a recent study, Abdous and He (2011) used text mining as a detection tool for the common technical problems faced by students taking video streaming courses.

Elsewhere, Zaiane and Luo (2001) analysed server logs to understand online learners’ behaviours in an effort to improve their web-based learning environments. Later, Zaiane (2002) used association rule mining to construct a recommender-system based on data from online learners’ profiles, access histories, and collective navigation patterns. This system can “intelligently” recommend learning activities or shortcuts to learners, based on the actions of previous learners. Similarly, Burr and Spennemann (2004) have pointed out that analysis of the patterns of user behaviour is important from both the technical and the pedagogical perspectives in order to predict network and traffic load, to align pedagogy with users’ behaviours, and to plan and deliver services in a timely manner.

For their part, Dringus and Ellis (2005) proposed a data mining approach for “discovering and building alternative representations for the data underlying asynchronous discussion forums.” This approach is intended to improve the instructor’s ability to evaluate the progress of a threaded discussion. More recently, Lin, Hsieh, and Chuang (2009) conducted a study to investigate the potential of an automatic genre classification system (GCS) that can be used to facilitate the coding process of the content analysis of a threaded discussion forum.

Of particular relevance to our study, we discovered several studies which have used various EDM techniques to predict students’ performance as measured by final grades. Minaei-Bidgoli and Punch (2003) used web-use features such correct answers, number of attempts for doing homework, total time spent on problems, participation in communication, and reading of material as predictors of students’ final grades. Their prediction accuracy varied between 51% and 86.8%, depending on the type of classifier used. Similarly, Falakmasir & Jafar (2010) used data mining to rank students’ activities which affected their performance, as measured by their final grade. Their findings suggest that students’ participation in virtual classrooms had the greatest impact on their final grades.

For their part, Zafra and Ventura (2009) used a grammar-guided genetic programming algorithm to predict students’ success or failure. These predictions were used to provide alternative learning activities that would enhance the students’ chances of success.

Using Learning Management Systems-generated student tracking data (Macfadyen & Dawson, 2010), we propose the development of a customizable dashboard-like reporting tool. This tool is intended to provide instructors with real-time data on both students’ engagement and the likelihood of their success. Unsurprisingly, their findings confirm that students’ contribution to the course discussion board is the strongest predictor of their success.

In reviewing the literature, Romero, Espejo, Zafra, Romero, and Ventura (2010) identified several avenues for using classification in educational settings: discovering student groups with similar characteristics, identifying learners with low motivations, proposing remedial actions, and predicting and classifying students using intelligent tutoring systems.

For their parts, Anand Kumar & Uma (2009) used the classification process to examine various attributes affecting student performance. Castellano and Martinez (2008) used collaborative filtering techniques to exploit students’ grades in order to generate group profiles which could facilitate academic orientation. Along the same lines, Vialardi et al. (2011) used data mining techniques which employed the students’ academic performance records to design a recommender system in support of the enrolment process.

In sum, this quick overview of the literature suggests that using various data mining techniques to predict students’ performance as measured by final grades has been examined by several different studies of traditional learning
management systems. However, none of the studies has explored the dynamics of online interaction in a live video streaming environment.

With these considerations in mind, we aim to apply both regression analysis and clustering analysis in order to explore students’ learning behaviours (students’ participation, login frequency, number of chat messages, and the type of questions submitted to instructor) along with their final grades. More specifically, we attempt to answer the following two questions:

- What are the major themes emerging from LVS students’ online questions?
- How do these emerging themes predict the students’ course grades?

![LVS interface](image)

**Figure 1. LVS interface**

**Methodology**

**Context of the study**

This study was conducted in a public research university in the mid-Atlantic region which serves 17,000 undergraduate and 6,000 graduate students and offers more than 70 bachelor’s degree programs, 60 master’s degree programs, and 35 doctoral degree programs in a variety of fields. Located in a major maritime, military, and commerce hub, this institution offers strong emphases in science, engineering, and technology, especially in the maritime and aerospace sciences. The university is also known as a national leader in technology-mediated distance learning, having served students at over 50 sites in Virginia, Arizona, and Washington state for more than twenty-five years. This extended distance learning capability provides the university with a variety of delivery mode options (i.e., ways in which a course can be delivered). Courses can be offered simultaneously via three different delivery formats: face-to-face, via satellite broadcasting, and via live video-streaming. Using the live video-streaming (LVS) delivery mode, students participate in the class, in real time, via personal computer, over which they view a live feed of the class lecture and during which they can interact with their instructor by sending text messages through the LVS course interface. Using the same interface, LVS students are able to chat with their LVS classmates during
class. At the receiving end (i.e., in the physical classroom), questions submitted by LVS students are displayed instantaneously on a monitor next to the instructor.

Instructors have the option to read/answer the messages, or to save, archive, and email them for later review. This tool is intended to enable instructors to seamlessly integrate LVS students into their classroom dynamic, without distraction and without overburdening instructors during their class time (Figure 1).

![Figure 2. Interaction in VS courses](image)

**Participants**

In total, 1,144 students completed 138 courses in a variety of subjects (e.g., accounting, computer engineering, information technology, human services, etc.) via the video streaming (VS) delivery mode during the Fall semester of 2009. All of the student-to-instructor questions, the two-way student-to-student chat messages, and the total login times were collected. Those VS students who never asked questions or chatted with their peers online were excluded from the actual analysis. The reasons why those students failed to get involved in the VS course discussion are suggested to be included in future investigation. One possible explanation is that some instructors never took the effort to invite their VS students to ask questions or to engage in online discussion. As a result of the pre-processing of the data, 298 students (those with complete information about their number of questions, number of chat messages, total login times, and final grade) were included in the data analysis. Due to factors such as privacy and university policy, the university’s registrar’s office could not provide us with the age or gender of these students, nor could we obtain the grading scales of each course. (The grading scale for each course at our university is determined by that course’s instructor.)

<table>
<thead>
<tr>
<th>Colleges</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Art and Letters</td>
<td>114 students (38.5%)</td>
</tr>
<tr>
<td>Education</td>
<td>79 students (26.5%)</td>
</tr>
<tr>
<td>Engineering</td>
<td>75 students (25%)</td>
</tr>
<tr>
<td>Science</td>
<td>23 students (7.7%)</td>
</tr>
<tr>
<td>Undecided</td>
<td>7 students (2.3%)</td>
</tr>
</tbody>
</table>

*Table 1. Distribution of students by college*

<table>
<thead>
<tr>
<th>Student academic level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate Students</td>
<td>138 students (46%)</td>
</tr>
<tr>
<td>Graduate Students</td>
<td>160 students (54%)</td>
</tr>
</tbody>
</table>

*Table 2. Distribution of students by academic level*

The questions and chat messages posted by those 298 students, along with their course ID, Student ID, the date, and a time stamp were saved in the database.
Clustering analysis: Identification of emerging themes

Our analytical approach included three phases: pre-processing, in which raw data is transformed into a usable format, mainly by cleaning, assigning attributes, and integrating data; mining the data by applying various mining strategies and tools such as classification, clustering, and visualization; and post-processing, which allows for interpretation and use of the gained knowledge in rethinking processes or in making decisions (Garcia et al., 2011).

All of the questions posted by the students were recorded in the Microsoft SQL server Database. To prepare the data processing for clustering analysis, we wrote a program using the PHP programming language to aggregate questions from the same students within the same course in order to form a case which included the sequence of questions posted by the students.

Subsequently, we used NVivo 9 software to apply an automatic coding technique to each of the student question cases. NVivo is a leading qualitative analysis tool on the market and has been used and tested by many researchers for content analysis (Zha, Kelly, Park & Fitzgerald, 2006). Automated coding one of NVivo 9’s features; it allows for automatic coding of a text document by text strings. After nodes were generated from each student question case, a clustering analysis was conducted in order to classify these nodes into different clusters with NVivo 9. According to NVivo, nodes are containers for specific themes, people, places, organizations, or other areas of interest.

Researchers can organize nodes into hierarchies – moving from more general topics (the parent node) to more specific topics (child nodes) – in order to support their particular research needs. Clustering analysis is a well-studied technique in data mining (Lin, et al., 2009) that uses an exploratory technique to visualize patterns by grouping sources which share similar words or attribute values, or which are coded similarly. From a data mining perspective, clustering is the unsupervised discovery of a hidden data concept. This approach is used in those situations in which a training set of pre-classified records is unavailable. In other words, this technique has the advantage of uncovering unanticipated trends, correlations, or patterns; no assumptions are made about the structure of the data (Chen & Liu, 2004).

The purpose of clustering analysis in this study is to classify students based on the student-shared characteristics in their questions. The cluster analysis tool in the NVivo 9 software confers upon researchers a different perspective on the unstructured textual data. Using the calculated similarity in each word that appears in the text of the nodes, NVivo 9 groups the nodes into a number of clusters. In our study, a statistical method named the Pearson correlation coefficient (\(-1 = \text{least similar}, 1 = \text{most similar}\)) was used as the similarity metric for the clustering analysis. The Pearson correlation coefficient is the preferred similarity metric used with NVivo. More information about the clustering analysis of NVivo can be found in NVivo’s online documentation website, [http://www.qsrinternational.com/support.aspx](http://www.qsrinternational.com/support.aspx).

To gain further insight from the textual questions or chat messages, we also applied the SPSS Clementine tool, which allowed us to analyse the unstructured textual data. The SPSS Clementine tool provides linguistic methods (extracting, grouping, indexing, etc.) for researchers to use in order to explore and extract key concepts from the text. As the result of the text mining, key concepts in our study were extracted and identified for analysis.

Measurement of final grade

The students’ final grades, submitted to the University Registrar by each course instructor, were supplied to us by the University Registrar. In the actual data analysis, the final grades were categorized into three groups: A- to A, B- to B+, and Others.

Quantitative data analysis: Predictive relationship between online question theme and final grade

In the current study, all of the quantitative data analysis was implemented using SPSS 17.0. Furthermore, the alpha levels were set at the .05 level for all significance tests.
Due to the ordinal nature of the final grade, ordinal logistic regression analysis (Norusis, 2008; O’Connell, 2006) was implemented in order to examine the predictive relationship between the online question theme as the predictor and the final grade as the criterion variable. Specifically, a cumulative odds model was fitted to the data. The use of ordinal logistic regression, which was closely related to logistic regression, helped to avoid the statistical consequences that could occur from the violation of assumptions in linear regression, such as normality of errors and linearity in the parameters (King, 2008). The log transformation in logistic regression also ensured that the predicted probabilities for the event of interest would range from 0 to 1 without imposing the numerical constraint on the predicted log odds from the logistic equation (Cohen, Cohen, West, & Aiken, 2003). Given the ordinal nature of the final grade, ordinal logistic regression was used to take into account the information regarding the rank ordering of the outcomes (Hosmer & Lemeshow, 2000).

The overall predictive utility of the ordinal logistic model with the online question theme as the predictor was assessed by testing the improvement of the model fit relative to the null model with no predictor, with the $\chi^2$ likelihood ratio test of the differences in deviances (O’Connell, 2006). The individual parameter estimate (i.e., the location coefficient) for the predictor variable was tested with the Wald test (Norusis, 2008). In ordinal logistic regression, two cutoffs (A- and B-) were used sequentially to form the cumulative odds equal to or higher than those two cutoffs, respectively. As a result, the probabilities of falling into three possible categories of final grade (A- to A; B- to B+; and Others) could be derived. Two different pseudo $R^2$ (Cox and Snell $R^2$ and Nagelkerke $R^2$) were also computed in order to quantify the overall model fit (O’Connell). The larger the pseudo $R^2$, the better the model fit.

The parallel lines assumption in ordinal logistic regression was checked with the $\chi^2$ likelihood ratio test (Norusis, 2008) to see if the relationship between those two research variables remained the same across two cutoffs (A- and B-).

Results

Identification of online question themes

In the current study, questions from each student during a semester were combined into one student entry so that students could be classified into different clusters based on the characteristics of their questions. The cluster analysis tool calculated each different word that appeared in the text of the entries by using the similarity metric. Then the entries were grouped into a number of clusters by NVivo 9, based on the calculated similarity index between each pair of entries. As a result, four major clusters of students were formed, based on the similarity of their questions. A multi-level, multiple cluster hierarchical structure was generated by clustering analysis (see Figure 3). These clusters were reviewed and interpreted collectively by two researchers and a graduate assistant. The two researchers had recently received specialized training about NVivo 9 from the software producer. Differences in the review were compared, discussed, and resolved to reach an agreement. The coding results were further reviewed and discussed with an educational researcher to validate their accuracy.

After a close review of the student questions in each cluster, four major themes were found (see Table 3).
Table 3. Four major themes in students’ online questions

<table>
<thead>
<tr>
<th>Cluster</th>
<th>N</th>
<th>Theme</th>
<th>Text Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90</td>
<td>Check-in</td>
<td>Class check-in</td>
</tr>
<tr>
<td>2</td>
<td>87</td>
<td>Deadline/Schedule</td>
<td>Submission deadline, exam schedule, lab schedule</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>Evaluation/Technical</td>
<td>Exam format, grading, office hours, and technical problems</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>Learning/Comprehension</td>
<td>Questions regarding course materials and assignments</td>
</tr>
</tbody>
</table>

Descriptive statistics of final grade

The descriptive statistics of students’ final grades by their online questions are listed in Table 4. Overall, about half (144, 48.32%) of the participants obtained a grade of A- or higher. Among the 298 participants, 90 posted mainly check-in questions and 87 posted questions related to deadline and schedule. The number of participants who posted questions mostly related to learning and comprehension was the lowest, relative to the number of their counterparts posting questions on other themes.

Table 4. Descriptive statistics of final grade by online question theme (N = 298)

<table>
<thead>
<tr>
<th>Final Grade</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>A- to A</td>
<td>43 47.78</td>
<td>36 41.38</td>
<td>32 45.71</td>
<td>33 64.71</td>
<td>144 48.32</td>
</tr>
<tr>
<td>B- to B+</td>
<td>28 31.11</td>
<td>29 33.33</td>
<td>23 32.86</td>
<td>15 29.41</td>
<td>95 31.88</td>
</tr>
<tr>
<td>Others</td>
<td>19 21.11</td>
<td>22 25.29</td>
<td>15 21.43</td>
<td>3 5.88</td>
<td>59 19.80</td>
</tr>
<tr>
<td>Total</td>
<td>90 100.00</td>
<td>87 100.00</td>
<td>70 100.00</td>
<td>51 100.00</td>
<td>298 100.00</td>
</tr>
</tbody>
</table>


Predictive relationship between online question theme and final grade

In the ordinal logistic regression model (see Table 5), the results of the chi-square likelihood ratio test supported a nonzero predictive relationship between the online question theme and the final grade, \( \chi^2 (3, N = 298) = 10.017, p < .05 \). Furthermore, the results did not indicate the violation of the parallel lines assumption, \( \chi^2 (3, N = 298) = 2.051, p > .05 \). Therefore, the predictive relationship between the online question theme and the final grade remained constant across two cutoffs of final grade (Norusis, 2008). The Cox and Snell \( R^2 \) and the Nagelkerke \( R^2 \) were .033 and .038 respectively, and indicated a modest predictive relationship. Overall, the online question theme would prove to be a useful predictor for the final grade.

The logistic regression coefficients (i.e., the location coefficients) for question themes 1, 2, and 3 were all positive and were statistically significant at the .05 level. Due to the way in which the ordinal logistic regression model was set up in SPSS (Norusis, 2008), the above statistically nonzero, positive regression coefficients suggested that the odds of getting a higher final grade, relative to all lower final grades at various cutoff values, were higher for the participants whose questions concerned learning/comprehension (Theme 4) in comparison with participants with the other three question themes (i.e., 1: Check-in; 2: Deadline/Schedule; 3: Evaluation/Technical). Specifically, for participants with the question theme of Learning/Comprehension, the odds of obtaining a grade equal to or higher than those two cutoffs (A- and B-), relative to all other lower grades, were 2.214 times higher than for the students whose questions had the theme of check-in, 3.020 times higher than those whose questions had the theme of deadline/schedule, and 2.361 times higher than the students whose questions had the theme of evaluation/technical. While using Question Theme 1, or Question Theme 2, or Question Theme 3 as the reference category respectively, no differences in the odds of obtaining better grades were found among the three theme groups.

The computed predicted probabilities of obtaining a final grade of A- to A+, B- to B+, and Others, respectively, for participants in those four question theme groups (1: Check-in; 2: Deadline/Schedule; 3: Evaluation/Technical; 4: Learning/Comprehension), were 47.07%, 32.89%, 20.04% in the Theme 1 group, 40.75%, 34.77%, 24.48% in the Theme 2 group, 45.48%, 33.43%, 21.09% in the Theme 3 group, and 66.31%, 23.51%, 10.17% in the Theme 4 group.
Table 5. Ordinal logistic model with online question theme as the predictor for final grade (N = 298)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question Theme 1</td>
<td>.795*</td>
<td>5.078</td>
</tr>
<tr>
<td>Question Theme 2</td>
<td>1.052*</td>
<td>8.834</td>
</tr>
<tr>
<td>Question Theme 3</td>
<td>.859*</td>
<td>5.452</td>
</tr>
<tr>
<td>Threshold</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade = A- to A</td>
<td>.677*</td>
<td>5.394</td>
</tr>
<tr>
<td>Grade = B- to B+</td>
<td>2.178*</td>
<td>47.867</td>
</tr>
</tbody>
</table>

Overall model evaluation

<table>
<thead>
<tr>
<th></th>
<th>( \chi^2 )</th>
<th>df</th>
<th>Cox and Snell ( R^2 )</th>
<th>Nagelkerke ( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood ratio test</td>
<td>10.017*</td>
<td>3</td>
<td>0.33</td>
<td>.38</td>
</tr>
</tbody>
</table>

Note. Question Theme 1: Check-in; Question Theme 2: Deadline/Schedule; Question Theme 3: Evaluation/Technical; Question Theme 4: Learning/Comprehension as the reference category.

Two cutoffs were set for the ordinal criterion variable, final grade, to examine how the increase in the faculty engagement score was related to the change in the odds, and in turn, to the probability of obtaining a higher final grade (O’Connell, 2006). The odds of obtaining a higher final grade at two cutoffs were the ratios of the probabilities of: A to all lower grades, and A through B- to all lower grades. The faculty engagement scores as the sample mean (i.e., 20.697 in raw score) and the one standard deviation (i.e., 5.560 in raw score) above the sample mean were examined to demonstrate the way in which the probability of obtaining a higher course final grade changed with the increase in faculty engagement (Norusis, 2008). Given an increase of one standard deviation in the faculty engagement score from the sample mean (i.e., from 20.697 to 26.257), the predicted probability of obtaining a final grade of A increased from 46.71% to 59.78% at the first cut-off. At the second cut-off, the predicted probability of obtaining a final grade of B- or higher increased from 78.79% to 86.30%.

Moreover, with the faculty engagement score as the sample mean (i.e., 20.697 in raw score), the predicted probabilities of obtaining one of those three categories of course final grade (A, A- to B-, or Other) were 46.71%, 32.08%, and 21.21% respectively. While the raw faculty engagement score increased by one standard deviation to 26.257, the predicted probabilities of obtaining one of those final grades became 59.78%, 26.52%, and 13.70% respectively. Therefore, the increase in the faculty engagement score was accompanied by the increased probability of obtaining a better course final grade.

Discussion

A student’s final grade depends on many factors, including the student’s motivation, learning style, and previous background, the instructor’s teaching and grading scales, the exam’s and assignment’s difficulty levels, etc. A holistic view of student demographic and institutional variables, as opposed to the single variable, must be examined in determining the overall online learning experience (Herbert, 2008).

In this study, our data shows that online VS student participation cannot be safely used to predict final grades. Perhaps the uniqueness of our VS interface (text-based chat in a live-video-streaming environment) explains our findings. Otherwise, previous studies including Macfadyen and Dawson’s study (2010) found that students’ participation and contribution to discussion boards in traditional learning management systems remain some of the strongest predictors of students’ success.

However, our analysis found that there is a correlation between questions posed to instructors and chat messages posted among students. Those who chat often also interact more often with their instructor.

We also analysed the chat messages (student-to-student communications) using the SPSS Clementine text mining tool. We noticed two outstanding concepts in the students’ chat messages (among themselves) and their frequency: they discussed technical problems (videos, sound, etc.) at 5% and test/exam issues at 2%. However, they addressed
the same concepts in their messages to the instructor with this frequency: technical problems at 2% and test/exam issues at 2%. Thus, it seems that students are more likely to discuss technology problems with their peers and try to help each other than to discuss those issues with their instructors.

The messages also revealed interaction patterns including topics related to project and assignment collaboration, discussion of grades, socialization, and greetings. In addition, the data reveals that students with a higher number of logins asked more questions and exchanged more chat messages with their classmates. In contrast, students with fewer logins rarely participated in the class; in fact, some of them rarely even logged into the system.

**Conclusions and future research**

This study was conducted in order to exploit the untapped data generated by LVS students. Our results revealed several student learning behaviours, ranging from active participation and interaction with the instructor to a lack of participation or even of attendance. Overall, our findings corroborate those of a previous study (Abdous & He, 2011). In spite of the limitations related to self-selection bias and to the use of final grades as a measurement of student learning outcomes (Abdous & Yen, 2010), we believe that we can provide some ways in which the learning experiences of LVS students can be improved and made more successful, based on our years of experience of working with faculty who teach VS courses. To this end, the following recommendations are made:

- Ensure faculty readiness and training prior to teaching LVS courses.
- Develop facilitation techniques to assist faculty in integrating LVS students into the dynamics of the classroom.
- Implement a tracking system for LVS students’ attendance.
- Encourage active participation and interaction during LVS sessions.
- Provide students with tips on effective participation and interaction during LVS sessions (writing messages, timing of questions, etc.)

As we make these recommendations, we reiterate that educational data-mining is clearly providing powerful analytical tools capable of converting untapped LMS and EPR data into critical decision-making information which has the capability of enhancing students’ learning experiences (Garcia et al., 2011). While adding to the body of literature, our hybrid approach provides a solid framework that can be used to exploit educational data to rethink and improve the learning experiences of students using some of the various new delivery modes that are currently reshaping higher education. Further understanding of students’ engagement and the dynamics of their interaction in and with these new delivery modes will contribute to the promulgation of an effective and engaging learning experience for all.

**References**


A Multidimensional Analysis Tool for Visualizing Online Interactions

Minjeong Kim1* and Eunchul Lee2

1Department of Teaching Education // 2Department of Education, Dankook University, Gyunggi-do, South Korea
448-701 // minjeong69@dankook.ac.kr // lec3918@gmail.com

*Corresponding author

ABSTRACT

This study proposes and verifies the performance of an analysis tool for visualizing online interactions. A review of the most widely used methods for analyzing online interactions, including quantitative analysis, content analysis, and social network analysis methods, indicates these analysis methods have some limitations resulting from their one-dimensional analysis approach. To overcome such limitations, we developed the Multidimensional Interaction Analysis Tool (MIAT) by considering the advantages of well-known methods and incorporating the concept of the comparative interaction level. To verify the performance of the MIAT as a tool for multidimensional visualization of online interactions, results of the one-dimensional interaction analyses and those of the MIAT were compared. Findings suggest that the MIAT can provide a more in-depth interpretation of online interaction than any one-dimensional analysis method. In addition, the MIAT allows researchers to customize their analysis frameworks based on their own theoretical backgrounds.

Keywords

Online Interaction, Visualization, Multidimensional Analysis Tool

Introduction

In recent years, problem-solving skills have attracted increasing attention from education researchers. The importance of collaborative learning, which facilitates literacy by enabling learners to cooperate with one another through various interactive activities, has been emphasized (An, Shin, & Lim, 2009; Pozzi, 2010). In particular, collaborative learning has been emphasized for implementation in online learning as well as face-to-face classes because, unlike face-to-face classes, online collaborative learning is available on an anywhere/anytime basis (Edge, 2006). In this regard, an increasing number of studies have focused on online collaborative learning, and some have explored factors that can facilitate online collaborative leaning (Benbunan-Fich & Hiltz, 1999; Sinclair, 2005; Yang, Newby, & Bill, 2008).

Some factors that have been found to promote online collaborative learning include individual characteristics of learners, instruction methods, the learning environment, learners’ motives, and type and level of online interaction. Researchers have suggested that, among the factors mentioned above, the type and level of online interaction are most likely to influence online collaborative learning (An et al., 2009; Daradoumis, Martínez-Mone, & Xhafa, 2006). For this reason, previous studies have typically focused on dynamic interaction in the online collaborative learning environment. In particular, a number of studies have attempted to find ways to analyze such interactions because different analysis methods tend to provide different information and interpretations.

Online interaction has attracted increasing attention from researchers and, thus, a number of studies have attempted to enhance existing analysis methods for measuring online interaction (Marra, 2006), including quantitative analysis, content analysis, and social network analysis (SNA) methods. The quantitative analysis method is used to investigate the level of online interaction by considering the number of posts by users, the number of replies, and the number of logins (Benbunan-Fich & Hiltz, 1999). A major advantage of this method is that it can easily quantify the level of online interaction. On the other hand, the content analysis method allows for an analysis of interaction types and levels by classifying learners’ posts based on certain criteria. Among various relationship analysis methods, the SNA method has recently been used by researchers to analyze the relationship among individuals within a certain group by treating those individuals as nodes and structuralizing message content into links (Hu & Racherla, 2008).

Although all of these methods are useful when analyzing online interaction, each of them cannot provide multidimensional aspects of online interaction, and thus each method focuses only on one aspect (e.g., the quantitative analysis method focuses on the number of interactions, the content analysis method on the types and levels of interaction, and the SNA method on the structure of the interaction). Researchers want a method that can provide rich information of online interpretation because they require an in-depth understanding of online interaction.
In addition, researchers want to visualize the analysis results of online interaction because visualization is a useful way to interpret complex interaction among group members more clearly (Hirschi, 2010). However, any principle or method analyzing interactions with a multidimensional approach has not been reported as of yet. Therefore, this study intends to develop an instrument that visualizes the results of the analysis on the basis of principles of multidimensional approaches to analyzing online interactions. A second objective of this study is to show how the results of multidimensional analysis and that of existing one-dimensional analyses are different.

Analysis methods for interactions in the online collaborative learning environment

Online interaction is defined as more than two people are giving and taking information to pursue their common learning goals in an online learning environment (Mäkitalo, Hääkkinen, Leinonen, & Jaervela, 2002). Online interaction includes relatively intensive information about the process of learners’ thinking and knowledge formation because it mostly happens in an asynchronous environment, which allows enough time for learners’ reflective thinking (An, Shin, & Lim, 2009; Blanchette, 2012; Garrison & Cleveland-Innes, 2005). To understand the online learning process properly, researchers must recognize that analyzing online interaction is an important issue. Depending on the types of analysis enacted, researchers can gather or lose valuable information (Blanchette, 2012). For this reason, many studies have addressed how to analyze online interaction. The following are representative ways to analyze online interaction.

Quantitative analysis of online interaction

The quantitative analysis method was the first method used for analyzing interactions in the online collaborative learning environment (Marra, 2006). This method considers the number of posts written and read, as well as replies and logins by learners. In addition, the quantitative analysis method also compares the points produced by adding the number of writings and readings for the level of online interaction (Benbunan-Fich & Hiltz, 1999; Gorghiua, Lindforstb, Gorghiuc, & Hämaläinend, 2011; Pozzi, 2010). For other methods, the average found from dividing the number of postings into the number of participants (Hewett, 2000), as well as a level of online interaction, was also analyzed by scoring the values of each message with certain criteria (Brooks & Jeong, 2006; Newman, Webb, & Cochrane, 1996).

This method typically employs relatively simple and objective quantitative data. Early studies considered this method to be the most objective for analyzing online interactions among learners since researchers were able to use diverse statistical methods based on quantitative data (Benbunan-Fich & Hiltz, 1999; Marra, 2006; Mason, 1992). However, this method is limited in that it provides only quantities for online interaction without analyzing the type and structure of online interaction or identifying important phenomena in the interaction process (Strijbos, Martens, Prins, & Jochems, 2006).

Content analysis of online interaction

The content analysis method is frequently used in research on online interaction because it can better analyze the type and level of online interaction than the quantitative analysis method, which allows only for limited information (George, 2008; Strijbos et al., 2006). The content analysis method characterizes the meaning of message content in a systematic and qualitative manner (George, 2008). The unit of analysis and the category of analysis play important roles in content analysis (Strijbos et al., 2006). The content of learners’ interactions is analyzed as messages, and such messages are classified based on the decided unit of analysis. A number of recent studies have used the content analysis method because of its ability to determine type, structure, and level of online interaction (Strijbos et al., 2006).

De Weber, Schellens, Valcke, & Van Keer (2006) described the framework of content analysis, and diverse analytical frameworks have been used for content analysis in the context of online collaborative learning. Among such frameworks, Henri’s (1992) framework has widely been used because its categories are clearly distinguished and its analysis method is relatively simple, allowing even non-experts to analyze messages exchanged during the learning process. Henri’s framework is composed of five categories: participative, social, interactive, cognitive, and...
metacognitive. Other widely-used frameworks are that by Gunawardena, Lowe, and Anderson (1997), which is composed of the following five categories: sharing/comparing information; discovery and exploration of dissonance; negotiation of meaning/co-construction of knowledge; testing and modification of proposed synthesis; and phrasing of agreement, statement, and application of the newly-constructed meaning, and that by Zhu's (1996), which is composed of the following six categories: answers, information sharing, discussion, comment, reflection, and scaffolding. These content analysis methods enhance the ability of researchers to gather a wide range of information from the interaction process, and thus the methods have been extensively used (Bassani, 2011; Kale, 2008). However, the content analysis method is limited in that, although it provides qualitative information about the types and levels of discourse content, it does not provide structural information of interaction.

Social network analysis of online interaction

The SNA method focuses on revealing the relationship and structure of online interactions among individuals in a group (Sternitzke, Bartkowski, & Schramm, 2009). The key advantage of the SNA method is its ability to visualize the relationship among individuals and the structure of their online interaction through nodes and links (Medina & Suthers, 2009; Sternitzke, et al., 2009). This method of analysis also provides information on personal contribution to interaction within the group (Contractor, Wasserman, & Faust, 2006), as well as varied information for analyzing interaction, such as its structure, flow, and processes. It is able to present results of the analysis after visualizing them (Bergs, 2006; Wasserman & Faust, 1989). In addition, SNA can visualize learning processes through group members’ interaction (Suthers & Rosen, 2011). Also, SNA provides quantitative data in the form of various indices, including centrality (the degree to which an individual occupies a central position in the network), concentration (the degree to which the entire network is concentrated toward the center), and density (the number of connections between individuals) (An et al., 2009; Heo, Lim, & Kim, 2010). However, although this method can be used to analyze the relationship and structure of online interactions among learners, it is limited in that specific types of messages cannot be analyzed.

Given the discussion above, each analysis method has beneficial aspects in terms of analyzing online interaction, but is limited in providing multiple aspects of online interaction due to its pursuit of one-dimensional analysis. In addition, the methods do not help researchers understand the results of interaction analysis more explicitly. Therefore, we suggest a multidimensional analysis method in order to overcome the limitations of one-dimensional analysis approaches.

Multidimensional Interaction Analysis Tool (MIAT)

We developed the Multidimensional Interaction Analysis Tool (MIAT), which can facilitate a multidimensional (quantitative analysis, content analysis, and structural analysis) analysis of online interactions among individuals in a group, as well as conceptualize those interactions in a visual way. In other words, the MIAT can simultaneously analyze the quantitative analysis/content analysis aspects of online interactions and the relationships among individuals in a group. In addition, the most remarkable feature of the MIAT is its ability to visualize all interactions among individuals in a certain group at a specific point in time. The principles of MIAT are as follows:

Unit of analysis

The unit of analysis is the most critical factor in content analysis (Woo & Reeves, 2007). In the MIAT, the unit of analysis is the message (the entire post under a title) because the MIAT considers the structure of the relationship among group members based on SNA.

Multidimensional principles of the MIAT

Quantitative analysis used in the MIAT carried two principles. One of them is using the frequency of a message, and the other is giving quantitative scores by evaluating the values of each message. For instance, assuming that there were three messages, and one contained false information whereas others contained exact information, the researcher
gave marks of less than three by deducting or subtracting points from the error message rather than giving a quantitative score of three with the number of messages. Thus, more plentiful information on interaction could be obtained if it was analyzed by giving quantitative scores from evaluating values of messages rather than simply using the frequency (Brooks & Jeong, 2006; Newman, et al., 1996). The criteria for assessing the values of each message in the MIAT can be freely selected by researchers. For instance, if a researcher desired to evaluate the values of messages by using the 10 criteria of Newman et al. (1996), he could give scores between -10–10 to each message. Additionally, if he desired to use criteria of Brooks & Jeong (2006), he could give marks between -1~1 by scoring +1 for a message that helped solve the assignment and -1 for a message that was not helpful. A key feature of the MIAT is it allows for criteria that assess values of each message to be adjustable by research goals or frameworks since online interactions can vary with learning contexts.

The MIAT used content analysis principles. Content analysis is a method that distinguishes message content by a certain category. The method can analyze the learning process shown during interactions and can achieve accurate, objective, consistent information on types and structures of interactions (De Weber, et al., 2006; Strijbos et al., 2006). The MIAT was developed for researchers to create their own frameworks for content analysis, in accordance with research purposes. For example, researchers can use Henri’s (1992) analyzing category or that of Zhu (1996) for MIAT analysis, with a category of content analysis according to the intention of the researcher. The MIAT also offers researchers the flexibility to input a self-generated category or use various other categories (see Figure 1). A lot of flexibility was given to selecting a category of content analysis because online interaction occurs in varied learning contexts, and the research frameworks of researchers desiring to analyze such interactions are also diverse.

![Figure 1. Creating the researcher’s analysis frameworks](image)

(e.g., using Henri’s (1992) categories (left) and Zhu’s (1996) categories (right))

![Figure 2. Basic model for analyzing the relationships among individuals](image)

The MIAT used a principle of online network analysis to examine a relationship between learners. In other words, if A posted something on a bulletin board and B made a B, and B influenced C. Wiki pages can be analyzed by using the history function as an editing method, which was different from the editing method of a bulletin board. Assuming that B modified A’s posting, and C either added details to this or asked a question about B’s modification, one could say that A influenced B, and B influenced C. Lastly, in the case of live chat, if A wrote a message and B wrote a message after A, and C wrote a message after B, one could say that A influenced B, and B influenced C (see Figure 2). The direction of arrows in the relationship model is the direction of message influence. Because B’s message comes after A’s, A’s message is expected to influence B, and thus the direction of the arrow goes from A to B. Therefore, the individual with the thickest outgoing arrows is expected to be the most influential group member, and the one with the thickest incoming arrows is most likely to be influenced by his interaction within the group. As
such, network analysis can clearly show the relationships among individuals in a group as well as the structure of their interactions. In this regard, the MIAT uses basic network analysis principles to analyze relationships among learners.

**Calculation of the level of online interaction**

To explain the principles for calculating a level of online interactions in the MIAT, we will use example data from a bulletin board system. First, criteria for assigning quantitative scores and categories for content analyses are required for explanations. Then, suppose the following four criteria of newness, importance, relevance, and accuracy are applied as a framework of quantitative analysis among 10 criteria used in research performed by Newman et al. (1996). Scores given ranged between 0 and 4 points. The remaining six criteria are excluded from this scoring since these are considered inappropriate because they are generally used to sort message content.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>New theme or details</th>
<th>Important information or details for completing the assignment</th>
<th>Relevant to the assignment or the discussion theme</th>
<th>Accurate Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now, suppose Henri’s (1992) framework is used in categories for content analysis. The original analysis categories of Henri include five types, but, for this example, only three categories are used, social, cognitive, and metacognitive, because a category analyzing the types of messages is required. Participative and interactive are the categories used to analyze a level and structure of interaction rather than the types of messages. Therefore, social, cognitive, and metacognitive were chosen for the analysis.

Principles that calculate a level of online interaction using the MIAT could be explained as follows. In order to calculate the level of interaction, the MIAT first requires the input of the analyzed data (the type of and score for each message) (see Figure 3). Then, the MIAT automatically provides a matrix of interaction scores (Figure 4).

![Figure 3. A sample screen showing input data for the MIAT](image)

The MIAT uses an interaction matrix to calculate two types of interaction levels: the total interaction level and the comparative interaction level. The total number of messages, average score, and standard deviation express the total interaction level. As shown in Figure 4, the total interaction level can be summarized as follows: the total number of messages=24, the total sum of scores per message=55, the average score=2.29, and the standard deviation=0.93. The MIAT uses the average score and the standard deviation to calculate the T-score for the comparative interaction level.
When using the sum or average of raw scores, it is difficult to identify comparative levels of interaction in a given group. Therefore, the MIAT uses the T–score (a standard score) to identify comparative scores for the group. For example, the cognitive interaction level between C and A is 3, which becomes 53.43 through the MIAT’s indexation (which uses the T–score). As a result, the cognitive interaction level between C and A is slightly higher than the average. The comparative interaction level is calculated as follows:

\[
T = (Z \times 10) + 50
\]

**Figure 4. Matrices of interaction scores**

**Interpretation of outputs**

We analyzed the data in 24 messages using the MIAT, and Figure 5 shows the results. In terms of the total interaction level, the total number of messages was 24, the average score was 2.29, and the standard deviation was 0.93. The comparative interaction level is indicated by the number next to each arrow; the larger the number, the higher the comparative interaction level. Further, the thicker the arrow, the larger the number of messages. The direction of the arrow indicates the direction of the interaction, and interaction types are classified into cognitive, metacognitive, and social categories (indicated by the style of the arrow).

As shown in Figure 5, Student A and Student C had a large number of links and thick arrows, which implies the interaction between these two students was the most active. Student A and Student B had fewer links than the other student pairs, and their arrows were thinner, indicating that the interaction between Student A and Student B was passive. However, the arrow for cognitive interaction (the solid arrow) of Student A and Student B was thicker than
those of the other student pairs. In addition, the active interaction between Student A and Student C was mainly social.

**Figure 5. The MIAT results**

**MIAT Implementation**

In this section, we will explain how the results of the MIAT analysis are different from those of other one-dimensional analysis methods through an example study. Since the level of participation in interaction is one of the important indicators of successful online collaborative learning, we conducted a study aiming to recognize group members’ interaction levels. Specifically, we wanted to identify the most active participant in a collaborative work. Through the study, we will try to explain how the results of the MIAT analysis are different from others.

**Participants and the task**

We conducted the MIAT analysis by considering a sample of 30 students taking an online course in education technology in the spring semester of 2011 at D University. The average age of these students was 21.7, and 67% were female. They had diverse majors, including human studies, social sciences, curriculum studies, engineering science, and art. The online collaborative task assigned to the class was based on instructional design. The students were randomly assigned to one of six groups (five students per group). Each group was expected to determine the theme of the instructional design through asynchronized interactions on an online bulletin board and to follow the instructional design for two weeks. Only the team showing the most active interaction was selected for this case analysis.

**Comparison analysis methods**

We compared the results from the MIAT with those from existing one-dimensional analysis methods (i.e., quantitative analysis, content analysis, and SNA methods).
For the quantitative analysis method, we used the method of Gorghiua et al. (2011), which is one of the most commonly used quantitative analysis methods of interaction levels. We counted the number of messages and the hit number of those messages.

For the content analysis method, we only used three categories (social, cognitive, and metacognitive) among Henri’s five categories (1992) because the other two categories (participative and interactive) do not pertain to the type of interaction. We classified each message based on these three categories and analyzed the interaction level based on the number of messages in each category.

We used NetMiner 2.4 for SNA. We analyzed the relationship among students by considering centrality, cohesion, and the number of messages sent and received.

Data analysis

The unit of analysis was the message. For the quantitative analysis method, we calculated the total number of posts as well as the number of hits. For the content analysis method, we analyzed the content of messages after their classification. Cohen’s kappa for the inter-rater reliability was 0.90. For SNA, we decided the direction of messages and calculated the number of messages. Cohen’s kappa for the inter-rater reliability was 0.92.

For the MIAT, we scored each message’s value based on the criteria in Table 1. Then, we categorized the message and decided its direction. Cohen’s kappa for the inter-rater reliability was 0.84 for scoring values, 0.93 for categorizing messages, and 0.92 for direction of messages. For inconsistent results, we reached an agreement through face-to-face discussions.

Analysis results of one-dimensional approaches

Table 2 shows the results of the quantitative analysis method. According to results from the quantitative analysis method, Student D produced the most posts and Student C yielded the most hits. Therefore, we can infer Student C and Student D were the most active participants in the group work.

<table>
<thead>
<tr>
<th></th>
<th>Posts</th>
<th>Hits</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23</td>
<td>171</td>
<td>194</td>
</tr>
<tr>
<td>B</td>
<td>19</td>
<td>114</td>
<td>133</td>
</tr>
<tr>
<td>C</td>
<td>23</td>
<td>212</td>
<td>235</td>
</tr>
<tr>
<td>D</td>
<td>27</td>
<td>175</td>
<td>202</td>
</tr>
<tr>
<td>E</td>
<td>22</td>
<td>170</td>
<td>192</td>
</tr>
<tr>
<td>Total number</td>
<td>114</td>
<td>842</td>
<td>956</td>
</tr>
</tbody>
</table>

According to results obtained using Henri’s (1992) framework, the most active participant is Student D with the largest sum. Among the Student D’s messages, the most common category was social. The next active participants were Student C and Student A, but they show different participation patterns. Student C’s messages are evenly distributed across all categories, but Student A’s messages were concentrated on social messages. Thus, content analysis gives information on the types of interaction as well as similar results of quantitative analysis.

Table 4. Results from the Content Analysis Method

<table>
<thead>
<tr>
<th></th>
<th>Cognitive</th>
<th>Metacognitive</th>
<th>Social</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>1</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>D</td>
<td>11</td>
<td>3</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>E</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>Total number</td>
<td>41</td>
<td>20</td>
<td>53</td>
<td>114</td>
</tr>
</tbody>
</table>
According to the results of the SNA, the total number of nodes was five and the total number of links was 20. The betweenness centrality of all nodes was 0 (see Figure 6), and the betweenness centrality stood for a degree of a node mediating the connection of other nodes. These results indicated that flow and exchange of information was even without a special focus on a certain student (Cho, Gay, Davidson, & Ingraffea, 2007). In terms of the cohesion analysis which is about an attractive force between nodes, there were five nodes and five clusters. This implies that no specific nodes gathered to form a cluster. Indeed, if nodes gather to form a cluster, the nodes in the cluster interact only with one another and not with nodes outside of the cluster. Therefore, the equal number of nodes and clusters indicated students engaged in balanced interactions. In terms of messages sent and received between nodes, Student C sent the highest number of messages, whereas Student D received the highest number of messages. This indicated that Student C interacted with other students most actively.

![Figure 6. SNA results](image)

**The results of MIAT method**

Figure 7 shows the results from the MIAT. Student C and Student E engaged in all three types of interactions (cognitive, metacognitive, and social). The comparative interaction level for cognitive interactions was 163.75 for Student C to Student E, and 79.96 for Student E to Student C. The comparative interaction level for metacognitive interactions was 121.36 for Student C to Student E, and 90.43 for Student E to Student C. The comparative interaction levels for cognitive and metacognitive interactions exceeded the average of 50, indicating the interaction between Student C and Student E contributed to their collaborative task.

**Comparison of analysis results**

The results of one-dimensional analyses (quantitative analysis, content analysis, and SNA) indicated that Student C and Student D are most active participants in collaborative group work. However, the MIAT showed slightly different results. According to the MIAT analysis, the most active participants were Student C and Student E. Student C was commonly identified as one of the most active participants, whereas Student E only appeared active in the results of the MIAT analysis. Therefore, who is the more active participant among collaborative work between Student D and Student E? According to the results of the one-dimensional analysis, Student D is a more active participant than Student E. Also, the results of quantitative analysis and SNA indicated Student D’s number of interactions is more than that of Student E. Additionally, the content analysis also indicated Student D’s number of
cognitive messages exceeded those of Student E. However, the results of the MIAT analysis indicated Student D’s T-score of cognitive messages was 598.58, which is lesser than Student E’s T-score of 622.77. In addition, Student D’s T-score of meta-cognitive messages was 107.61, which was also below Student E’s T-score of 382.27.

The variability in the results is due to differences between one-dimensional analyses and the MIAT analysis. One-dimensional analyses simply count the number of messages or the types of messages, whereas the MIAT considers the type of interaction and the comparative interaction level. For example, there was almost no difference in the total number of messages between Student A and Student C, and between Student C and Student E. In this case, results of the SNA indicated balanced interactions among group members. However, results from the MIAT convey a different story. Student A and Student C were most likely to engage in social interactions, whereas Student C and Student E were most likely to engage in cognitive interactions. These results indicated interactions between Students C and E were more likely to contribute to collaborative work than were the interactions between Students A and C. Previous studies have found that cognitive interactions directly influence the problem-solving activity of learners (Veerman & Veldhuis-Diermanse, 2001).

In sum, the MIAT considers both the type of interaction and the comparative interaction level in order to provide more specific and in-depth interpretation information on interactions among group members. However, SNA considers only the number of messages and the direction of messages. The results obtained through an analysis were differentiated by whether the approach was one-dimensional or multidimensional. Results showed the multidimensional approach can provide more in-depth information about the learning process and the structure of online interactions. This is because interactions could be understood more deeply when taking the multidimensional approach into account, compared to working only with a one-dimensional approach (Tomsic & Suthers, 2005).

Figure 7. The MIAT results
Conclusion and implications

Existing quantitative or content analysis methods for analyzing interactions among group members are limited in that they have difficulty providing rich information on such interactions. This is because such methods take a one-dimensional approach. For instance, Driver (2002) and Chiu and Hsiao (2010) examined the effects of group size in online collaborative learning but provided different findings. Driver (2002) found no differences in interactions among group members when comparing large and small groups, whereas Chiu and Hsiao (2010) concluded that interactions among small-group members were more effective than those among large-group members. The researchers obtained different results because they used different methods. Driver used a self-reported questionnaire to measure interactions among students, whereas Chiu and Hsiao conducted a content analysis. This example suggests a need for caution when interpreting interaction results obtained using a method based on a one-dimensional approach because the results can vary according to the method used. Methods based on a one-dimensional approach have difficulty providing sufficient information for an in-depth analysis of the interactions among group members. Thus, this research was performed based upon the premise that multidimensional analysis was pivotal for an in-depth understanding on interactions.

Looking into the methods for analyzing interactions of online cooperative activities used until now, we came across a case using either one of quantitative analysis, content analysis, and relational methods, as well as a case using two analyses in tandem (e.g., Newman et al., 1996; Tomsic & Suthers, 2005). However, a multidimensional analysis considering all types of analyses simultaneously has rarely been performed. Multidimensional analysis could provide information on teaching and learning to instructors conducting online lessons since it analyzes the details of interactions and provides visual information on a learner’s structural relationships in an online cooperative study. Therefore, this study introduced the principles of developing the MIAT, a multidimensional instrument analyzing online interactions and considering quantitative analysis, content analysis, and relational analyses simultaneously. We also explained its advantages in analyzing online interactions by comparing its performance with existing analysis methods.

Utilities in the MIAT, in comparison with existing methods for analyzing online interactions, could be arranged as follows. The MIAT could provide the results of quantitative analysis, content analysis, and relational analyses simultaneously since the functions of one-dimensional analytical methods were integrated for the MIAT. It also provided visual results of an analysis so researchers could see the flow and processes of interactions as well as the visual, relational structure between learners. Numerous scholars have acknowledged that visualization is an effective way to support a deep understanding of interactions (Medina & Suthers, 2009; Saltz, Roxanne, & Turoff, 2004).

Also, the MIAT provided flexibility, which could modify a specific analytical framework of existing analyzing methods into various forms. For instance, some researchers might desire to use Henri’s framework for a content analytical framework for interaction, while other researchers might wish to use the framework of Gunawardena and his colleagues (1997). The MIAT uses a content analytical framework, where an appropriate framework could be entered directly in accordance with the researcher’s purpose or background of studies. Moreover, quantitative scoring criteria for assessing the values of each message could be entered by a researcher’s intention or framework of studies (Newman and his colleagues’ standard was used as a criterion for assigning quantitative scores and Henri’s framework was used for content analysis in an example analysis in this study). It is expected that the MIAT can provide meaningful information for instructors or researchers due to its characteristics of flexible analysis framework and visualization. The information provided by the MIAT would vary and be more robust than information provided by one-dimensional analysis methods of interactions. De Weber et al. (2006) indicated that coding categories for interaction are developed to analyze the process of knowledge acquisition, sharing, and formation. This means the information from interaction analysis pertains to the learning process. Thus, the MIAT would provide some useful information about the learning process by providing multidimensional information.

Limitations and future study

The purpose of this study was to provide instructors and researchers with more teaching-learning information through analyzing interactions. This was accomplished by developing an instrument that suggested principles for the multidimensional approach to online interaction analysis. However, there were a few limitations to the study. First, the MIAT confined the scope of analysis in analyzing online interactions. Relational analysis was one of the most
important analytic functions for the MIAT, so the context must be definite for the utterance. In this case, it is considered as interaction where an individual wrote after another individual. Although the writing targeted unspecified individuals, it is simply considered an interaction between previous writer and next writer. So in this case, the MIAT cannot analyze the interaction completely.

Second, the MIAT focused on quantitative analysis, content analysis, and relational analysis. However, the MIAT did not provide information on a change of interactions according to time. The MIAT would function as a more powerful analytic instrument if it also considered a change of learners, in accordance with time, while learning for a certain period of time.

Third, the MIAT allowed for an analytic framework in which a researcher was interested in using a standard for quantitative and content analyses. The MIAT also included flexibility, as it could select the number of people for a study team according to the intentions of the researcher. Nevertheless, the researcher must enter scores assigned to messages for content analysis and evaluation during this process. Therefore, the MIAT’s capabilities could be optimized to the researcher’s purpose. However, this requires the user’s efforts and the MIAT is not completely automated.

Finally, a few follow-up studies are suggested to improve meaningful use of the MIAT. This study focused on the necessity of multidimensional analysis for online interactions, as well as developing an instrument for such a purpose. However, a study that applies the developed MIAT in a research context, with a researcher’s theoretical framework, should be performed as well. Also, this study analyzed interactions on ordinary bulletin boards in order to examine the relative advantages and disadvantages offered by the MIAT over existing analyzing methods. It would be beneficial to investigate further what kind of analytic information can be provided by the MIAT in various cooperative activities online (e.g., wiki, live chat).

Furthermore, this study focused on comparing the MIAT with existing analysis methods to verify whether the actual results of the MIAT analysis were appropriate and not supplemented with content analysis research. This concern must also be addressed by additional studies. The process requires future work to ensure the results of the MIAT analysis are valid. This should involve additional information, such as learner interviews, to analyze online interactions more precisely and make appropriate conclusions.

Acknowledgements

The present research was conducted by the research fund of Dankook University in 2011.

References


Teaching Analytics: A Clustering and Triangulation Study of Digital Library User Data

Beijie Xu and Mimi Recker*
Department of Instructional Technology & Learning Sciences, Utah State University, Logan, UT, U.S.A. // beijie.xu@aggiemail.usu.edu // mimi.recker@usu.edu

*Corresponding author

ABSTRACT

Teachers and students increasingly enjoy unprecedented access to abundant web resources and digital libraries to enhance and enrich their classroom experiences. However, due to the distributed nature of such systems, conventional educational research methods, such as surveys and observations, provide only limited snapshots. In addition, educational data mining, as an emergent research approach, has seldom been used to explore teachers’ online behaviors when using digital libraries. Building upon results from a preliminary study, this article presents results from a clustering study of teachers’ usage patterns while using an educational digital library tool, called the Instructional Architect. The clustering approach employed a robust statistical model called latent class analysis. In addition, frequent itemsets mining was used to clean and extract common patterns from the clusters initially generated. The final clusters identified three groups of teachers in the IA: key brokers, insular classroom practitioners, and inactive islanders. Identified clusters were triangulated with data collected in teachers’ registration profiles. Results showed that increased teaching experience and comfort with technology were related to teachers’ effectiveness in using the IA.

Keywords

Educational data mining, Latent class analysis, Teacher usage patterns

Introduction

Increasingly, education and training are delivered beyond the constraints of the classroom environment, and the increasingly widespread availability of online repositories, educational digital libraries, and their associated tools are major catalysts for these changes (Borgman et al., 2008; Choudhury, Hobbs, & Lorie, 2002). Teachers, of course, are a primary intended audience of educational digital libraries. Studies have shown that teachers use digital libraries and web resources in many ways, including lesson planning, curriculum planning (Carlson & Reidy, 2004; Perrault, 2007; Sumner & CCS Team, 2010), and looking for examples, activities as well as illustrations to complement textbook materials (Barker, 2009; Sumner & CCS Team, 2010; Tanni, 2008). Less frequently mentioned ways are learning about teaching areas (Sumner & CCS Team, 2010; Tanni, 2008), networking to find out what other teachers do (Recker, 2006), and conducting research (Recker et al., 2007). These studies, however, were generally conducted in laboratory-like settings, using traditional research methods, such as interview, survey, and observation.

Due to the distributed nature of the Web, traditional research methods and data sources do not support a thorough understanding of teachers’ online behaviors in large online repositories. In response, web-based educational applications are increasingly engineered to capture users’ fine-grained behaviors in real-time, and thus provide an exciting opportunity for researchers to analyze these massive datasets, and hence better understand online users (Romero & Ventura, 2007).

These records of access patterns can provide an overall picture of digital library users and their usage behaviors. With the help of modern data mining techniques—the discovery and extraction of implicit knowledge from one or more large databases (Han & Kamber, 2006; Pahl & Donnellan, 2002; Romero & Ventura, 2007)—the data can further be analyzed to gain an even deeper understanding of users. Yet, despite the wealth of fine-grained usage data, data mining has seldom been applied to digital library user datasets, especially when studying teacher users.

The study reported in this article used a particular digital library tool, called the Instructional Architect (IA.usu.edu), which supports teachers in authoring and sharing instructional activities using online resources (Recker, 2006). The IA was used as a test bed for investigating how the data mining process in general, and clustering methods in particular, can help identify the different and diverse teacher groups based on their online usage patterns. This study built substantially on results from a preliminary study that also used a clustering approach (Xu & Recker, in press). In particular, both studies relied on a clustering approach that used a robust statistical model, latent class analysis.
(LCA). In addition, this study used more refined user feature space, and frequent itemsets mining was used to clean and extract common patterns from the clusters initially generated. Lastly, as a means of validation the clustering results, we explored the relationship between teachers’ characteristics (comfort level with technology and teaching experience) and the teacher clusters that emerged from the study.

This article is organized as follows. The literature review first describes the Knowledge Discovery and Data Mining (KDD) process, and several clustering studies conducted with educational datasets. This is followed by a brief introduction to the Instructional Architect tool. We then describe our data mining approach, starting from data collection and selection, through data analysis, interpretation, and inference. Finally, as part of the interpretation process, we triangulated data from teachers’ registration profiles to validate the clustering results. We conclude with the implications, contributions, and limitations of this work.

This section describes the general data mining approach, and reviews several clustering studies set within educational contexts.

Educational data mining

There is increasing interest in applying data mining (DM) to the evaluation of web-based educational systems, making educational data mining (EDM) a rising and promising research field (Romero & Ventura, 2007). Data mining is the discovery and extraction of implicit knowledge from one or more large databases, data warehouses, and other massive information repositories (Han & Kamber, 2006; Pahl & Donnellan, 2002; Romero & Ventura, 2007). When the context is the Web, it is sometimes explicitly termed web mining (Cooley, Mobasher, & Srivastava, 1997). Educational data mining, as an emerging discipline, is concerned with applying data mining methods for exploring unique types of data that come from educational settings (Baker & Yacef, 2009). As web-based educational applications are able to record users’ fine-grained behaviors in real-time, a massive amount of data becomes available for researchers to analyze in order to better understand an application’s impact, usage, and users (Romero & Ventura, 2007).

The knowledge discovery and data mining (KDD) process typically consists of three phases: 1) preprocessing datasets, 2) applying data mining algorithms to analyze the data, and 3) post-processing results (Cooley et al., 1997; Romero & Ventura, 2007). Data preprocessing refers to all the steps necessary to convert a raw dataset to a form that can be ingested into a data mining algorithm. It may include any of the following tasks: data cleaning, missing value imputation, data transformation, and data integration. The application of data mining algorithms usually has one of two purposes: description and prediction. Description aims at finding human-interpretable patterns to describe the data; prediction attempts to discover relationships between variables, in order to predict the unknown or future values of similar variables. Currently, there is no universal standard for post-processing and evaluating data mining results. Typical interpretation techniques draw from a number of fields such as statistics, data visualization, and usability studies.

Clustering studies in educational settings

The increasing availability of educational datasets and the evolution of data mining algorithms have made educational data mining a major interdisciplinary area, lying between the fields of education and information/computer sciences. Based on Romero and Ventura’s (2007) educational data mining survey, most commonly used data mining techniques include statistical data mining, classification, clustering, association rule mining, and sequential pattern mining. This study focused on using clustering approach to analyze teachers’ online behaviors when using a digital library tool. As such, several clustering studies using in educational datasets are reviewed.

Hübscher, Puntambekar, & Nye (2007) used K-means and hierarchical clustering techniques to group students who used CoMPASS, an educational hypermedia system that helps students understand relationships between science concepts and principles. K-means is a clustering analysis method that aims to partition \( n \) data points into \( k \) clusters in which each data point belongs to the cluster with the nearest cluster center. Hierarchical clustering is a clustering analysis method that seeks to build a hierarchy of clusters. In CoMPASS, navigation data was collected in the form
of navigation events, where each event consisted of a timestamp, a student name, and a science concept. After preprocessing, K-means and hierarchical clustering algorithms were used to find student clusters based on the structural similarity between navigation matrices.

Durfee, Schneberger, & Amoroso (2007) analyzed the relationship between student characteristics and their adoption and use of particular computer-based training software, using factor analysis and self-organizing map (SOM) techniques. Survey responses to questions regarding user demographics, computer skills, and experience with the software were collected from over 40 undergraduate students. They used SOM to cluster and visualize the dataset. By visually analyzing the similarity and difference of the shades and borders, four resulting student clusters were identified. Finally, a t-test on performance scores supported the clustering decisions.

Wang, Weng, Su, & Tseng (2004) combined sequential pattern mining with a clustering algorithm to study students’ learning portfolios. The authors first defined each student’s sequence of learning activities as a learning sequences, \( LS = \langle s_1, s_2, \ldots, s_n \rangle \), where \( s_i \) was a content block. They then applied a sequential pattern mining algorithm to find the set of maximal frequent learning patterns from learning sequences. The discovered patterns were considered as variables in a feature vector. For each learner, the value of bit \( i \) was set as 1 if the pattern \( i \) was a subsequent of the original learning sequence, 0 otherwise. After the feature vectors were extracted, a clustering algorithm called ISODATA was used to group users into four clusters.

The literature review only identified one clustering study investigating teachers’ use of an educational digital library tool. In this study, a clustering approach was applied to model and discover patterns in teachers’ using an online curriculum planner (Maull, Saldivar, & Sumner, 2010). In this study, user sessions were first abstracted, and 27 features were selected for clustering experiments. The study then used K-means and expectation-maximum (EM) likelihood to cluster the user sessions. The two algorithms identified very similar patterns in the largest clusters, such as clicking on instructional support materials, embedded assessments, and answers and teaching tips. However, the authors acknowledged that their study was preliminary, in that there was not complete agreement between the different algorithms on top cluster features or cluster sizes.

There are other clustering studies documented in the literature on educational web mining, however, the above examples are sufficient in revealing some major considerations in discovering user groups in the context of online environments, as follows:

- A user-model must be carefully defined that accounts for the task and domain. Navigational paths, online performance, user characteristics, and a user’s prior knowledge are all good candidates for user features.
- Clustering is a generic definition for a certain type of data mining method. Researchers must select the clustering algorithm appropriate for their studies; however, different approaches may produce different results.
- Other data mining methods such as rule discovery, dimensionality reduction, and filling in missing values can be used with clustering algorithms to achieve a better grouping effect.
- To better understand online user behaviors and produce more useful information, the data mining results should be used in conjunction with other data.
- As an indispensable component of the KDD process, evaluation of the clustering results should be conducted if at all possible.

**Teachers’ use of digital libraries**

As noted, the research context is teachers’ use of digital libraries, an area that is seeing explosive growth in educational settings (Borgman et al., 2008). While prior work has examined the influence of teacher characteristics (such as teaching experience, information literacy skills, and usage patterns), little work has identified quantitative evidence linking these.

For example, prior work has noted that teachers often lack the necessary information seeking and integration skills to effectively use online resources (Perrault, 2007; Tanni, 2008). In a nation-wide survey on teachers’ perceived value of the Internet, Barker (2009) found a positive correlation between teacher self-reports of the perceived value of the Internet in teaching, and use of hardware/electronic media. However, this work failed to find any correlation between teachers’ perceptions and years of teaching experience.
To examine usage, researchers are increasingly turning to web metrics, a close kin to the EDM family. In a review of four educational digital libraries projects, Khoo et al. (2008) reviewed the use and utility of web metrics. Others have examined such metrics in conjunction with other sources of data, thereby seeking triangulation and complementarity in findings (Greene, Caracelli, & Graham, 1989). In an evaluation of a digital library service, the Curriculum Customization Service (CCS), Sumner & CCS Team (2010) reported interview data of middle and high school science teachers, and examined how their experiences were supported and clarified by usage log data. However, web metrics do not always agree with teachers’ own stories. For example, in Shreeves and Kirkham’s (2004) usability testing of a search portal, 65% of the users reported using the advanced search features; however, transaction log analyses did not support these claims. As such, these studies raise important questions. Since every research method has limitations, which should be trusted when there are discrepancies? Can data triangulation be conducted to help resolve these discrepancies?

**Technology context: The instructional architect**

This research is set within the context of the Instructional Architect (IA.usu.edu), a lightweight, web-based tool developed for supporting authoring of simple instructional activities using online learning resources in the National Science Digital Library (NSDL.org) and on the Web (Recker, 2006). With the IA, teachers are able to search for, select, sequence, annotate, and reuse online learning resources to create instructional web pages, called IA projects. These IA projects (or, projects, for short) can be kept private (private-view), made available to only students (student-view), or to the wider Web (public-view). Anyone can visit a public-view IA project, students can access their teachers’ student-view IA projects through their student accounts, and private IA projects are only viewable by the author. Any registered teacher can make a duplicate of any public IA project by clicking the copy button at the bottom of the project. In this way, the IA provides a service level for supporting a teacher community around creating and sharing instructional resources and activities. To date, the IA has over 7,000 registered users who have created over 16,000 IA projects.

To use the IA, a teacher must first register by creating a free IA account, which provides exclusive access to his/her saved resources and projects. As part of the registration process, teachers were asked two optional profile questions: years of teaching experience and comfort level with technology.

After logging in, the IA offers two major usage modes: **resource management** and **project management**. In the resource management mode, teachers can search for and store links to NSDL resources, web resources, as well as to other users’ IA projects. These links are added to teachers’ personal collections within the IA. Within the IA’s project management interface, teachers only need to enter the IA project’s title, overview, and content for the IA system to dynamically generate a webpage which can then be published. Figure 1 shows an example of a teacher-created IA project.

**Purpose and research questions**

As noted above, this study relied on results from a preliminary study organized around the KDD process and using **latent class analysis** (described below) as the clustering algorithm with the same usage data (Xu & Recker, in press). Preliminary results demonstrated LCA’s utility by clustering teachers into seven groups based on thirteen features drawn from teachers’ online behaviors. Results, however, also suggested the following improvements: 1) a more parsimonious user feature space, 2) inclusion of a clustering pruning process to make the clustering results less ambiguous, and 3) validation of clustering results by triangulating with teacher profile data.

As such, the purpose of this study is to build upon results from the preliminary study to better understand teachers’ use of the IA. In particular, by implementing the suggested improvements, what usage patterns and clusters emerge when mining teacher usage data? What inferences can be made about teachers’ behaviors from the discovered usage patterns? Finally, how can user patterns be combined with more traditional user data for triangulation purposes?
Results

Phase 1 -- Data preprocessing: Generating the user feature space

The dataset included usage data from 661 teachers who registered in the IA in 2009 and had created either public-view or student-view project(s) (57% of the 1,164 teachers who registered during that period). As outlined above, a teacher can assume three general roles in the IA environment: project authoring, project usage, and navigation. In the preliminary study, we generated an initial list of 13 indicators based on teachers’ possible behaviors in each of these three roles (Xu & Recker, in press). Clustering results from this preliminary study were used to inform how we reduced the complexity of the feature space, by fine-tuning or removing some indicators (see Table 1). Note that the number of student visits referred to the number of times a teacher’s project was viewed by his/her students. The number of peer visits referred to the number of times a teacher’s projects was viewed by other IA users.

Our dataset also contained variables that were rather skewed or had outliers. The presence of outliers can lead to inflated variance and error rate, as well as distorted estimation of parameters in statistical models (Zimmerman, 1994). For example, 98% users’ projects had less than 150 maximum number of student visits; the inclusion of the 2% users with more than 150 maximum number of student visits increased the mean value by 2.5 times (from 4.29 to 10.96) and the standard deviation by almost 4.5 times (from 12.58 to 56.48). Thus, eight features in the original dataset were scaled into three levels using ordinal variables. The remaining feature, number of projects, was segmented into two levels. Generally, equal intervals were used to discretize a continuous variable, except for those features with extremely skewed distributions. Then, professional opinion influenced the segmentation process.
Phase 2 -- Applying data mining algorithms

This study also used Latent Class Analysis (LCA) (Magidson & Vermunt, 2004) to classify registered teacher users into groups. LCA is a model-based cluster analysis technique in that a statistical model (a mixture of probability distributions) is postulated for the population based on a set of sample data. LCA offers several advantages over traditional clustering approaches such as K-means: 1) for each data point, it assigns a probability to the cluster membership, instead of relying on the distances to cluster means; 2) it provides various diagnostics such as common statistics, Log-likelihood (LL), Bayesian information criterion (BIC) and $p$-value to determine the number of clusters and the significance of variables’ effects; 3) it accepts variables of mixed types without the need to standardize or normalize them; and 4) it allows for the inclusion of demographic and other exogenous variables either as active or inactive factors (Magidson & Vermunt, 2004).

The traditional LCA (Goodman, 1974) assumes that each observation belongs to only one of the $K$ latent classes, and that all the manifest variables are locally independent of each other. Local dependence means that all associations among the variables are solely explained by the latent classes; there are no external associations between any pair of input variables. An example of an external association is having two survey items with similar wording in the questions (Magidson & Vermunt, 2004).

LCA uses the maximum likelihood method for parameter estimation. It starts with an expectation-maximization (EM) algorithm and then switches to the Newton-Raphson algorithm when it is close enough to the final solution. In this way, the advantages of both algorithms, the stability of EM and the speed of Newton-Raphson when it is close to the optimum solution, are exploited.

<table>
<thead>
<tr>
<th>Role</th>
<th>Indicators</th>
<th>Raw data</th>
<th>Segmentation</th>
<th>Range of original values</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project authoring</td>
<td>Number of projects</td>
<td>Number of projects</td>
<td>1</td>
<td>1</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2 $\sim$ 10</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td>Average number of resources per project</td>
<td>Project content</td>
<td>1</td>
<td>0 $\sim$ 2</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>3 $\sim$ 4</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>5 $\sim$ 44</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>Average number of words per project content</td>
<td>Project content</td>
<td>1</td>
<td>0 $\sim$ 32</td>
<td>222</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>33 $\sim$ 167</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>168 $\sim$ 2843</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>Average number of words per project overview</td>
<td>Project content</td>
<td>1</td>
<td>0 $\sim$ 11</td>
<td>247</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>12 $\sim$ 21</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>22 $\sim$ 293</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>Number of copied projects$^a$</td>
<td>Project originality</td>
<td>1</td>
<td>0</td>
<td>508</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2 $\sim$ 18</td>
<td>75</td>
</tr>
<tr>
<td>Project usage</td>
<td>Maximum number of peer visits</td>
<td>Project visits</td>
<td>1</td>
<td>0</td>
<td>343</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2 $\sim$ 164</td>
<td>193</td>
</tr>
<tr>
<td></td>
<td>Maximum number of student visits</td>
<td>Project visits</td>
<td>1</td>
<td>0</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1 $\sim$ 5</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>6 $\sim$ 1022</td>
<td>125</td>
</tr>
<tr>
<td>Navigation</td>
<td>Number of visits to the IA</td>
<td>Transaction data</td>
<td>1</td>
<td>1 $\sim$ 4</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>5 $\sim$ 8</td>
<td>217</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>9 $\sim$ 57</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td>Number of project browses</td>
<td>Transaction data</td>
<td>1</td>
<td>0</td>
<td>212</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1 $\sim$ 4</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>5 $\sim$ 134</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td>Number of copied projects$^a$</td>
<td>Project originality</td>
<td>1</td>
<td>0</td>
<td>508</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2 $\sim$ 18</td>
<td>75</td>
</tr>
</tbody>
</table>

$^a$The number of copied projects belong to both the project authoring and navigation role.
The next section describes how LCA was applied to the user feature space, and how the final user clusters were selected.

The user feature space (consisting of nine features in three roles) was used as the input for the LCA. Due to the unsupervised nature of clustering studies, it is hard to determine the number of clusters without any predefined guidelines. Therefore, we explored the clustering problem with different \( k \)'s, and then observed the common patterns emerging from different settings. By doing this, the clustering results as defined by common patterns were robust and not contingent on a particular setting. The data analysis consisted of four steps: (1) generating preliminary clusters, (2) deriving user patterns, (3) mining frequent user patterns, and finally (4) selecting the final user clusters. Step 1 was used to generate preliminary LCA models. Steps 2 through 4 were used to extract the common patterns, in other words, the final user clusters.

**Step 1: Generating preliminary clusters.** All LCA models were generated starting from the number of clusters \( k = 3 \) to \( k = 15 \). With all models, we monitored three criteria (\( R^2 \), BVR, BIC) to ensure that the optimal model could be achieved. \( R^2 \), also called the coefficient of determination, is the proportion of the total variation of scores from the grand mean that is accounted for by group membership (Aron, Aron, & Coup's, 2009; Howell, 2007). In terms of the LCA, it means how much of the variance of each indicator is explained by an LCA model (Statistical Innovations, 2005). If an indicator has a very small \( R^2 \) value, then it is making little contribution to current latent class analysis model, and the current model needs to be adjusted. Bivariate residual (BVR) in an LCA model is a local measure of model fit by assessing the extent to which the observed association between any pair of indicators is explained by a model (Statistical Innovation, 2005). If we encountered a BVR greater than 1 for any pair of indicators, we manually forced a correlation between them. BIC is a posterior estimation of model fit based on comparing probabilities that each of the models under consideration is the true model that generates the observed data (Kuha, 2004). A model with a lower BIC value is preferred over a model with a higher value. The BIC measure is widely used to help in LCA model selection.

The best LCA models under different number of clusters \( k \) were selected using the three measures described above. We found that some resulting clusters were too small to demonstrate a reliable pattern. For instance, some clusters only had 10 users, with several of their indicators distributed across all segmentation levels. This means that after filtering out the outliers, the few users left did not demonstrate a distinctive cluster-wise pattern. In order to obtain representative user patterns, these kinds of small-sized clusters were excluded and only clusters greater than a certain threshold, \( \alpha \), were used. \( \alpha \) was defined as the smaller of the two: 1) 10% of the total number of users, or 2) \( N / k \), where \( N \) was the total number of users and \( k \) was the cluster size. In the end, 59 clusters from models of different \( k \) were above their respective thresholds.

**Step 2: Deriving user patterns.** A valid cluster was then converted to a piece of user pattern, which was a conjunction of the themes of individual features within a cluster. As noted in Table 1, each feature was segmented to two or three levels. When deriving user patterns, an individual feature’s theme for a given cluster referred to how users within this cluster distributed among the levels of this particular feature. For example, the number of projects was the only two-level indicator, and it had two themes (one project, and more than one projects). All other indicators had three levels, and thus, in theory, could produce five themes: 1) the lowest level is dominant, 2) the lowest two levels are dominant, 3) the middle level is dominant, 4) the highest two levels are dominant, and 5) the highest level is dominant. To be a dominant level (e.g., the lowest level is dominant) or dominant adjacent levels (e.g., the lowest two levels are dominant), more than 70% users must fall into such level(s). For instance, when the number of clusters \( k = 3 \), 84.6% teachers in the 2nd cluster had only a few words (the lowest level) for their project content, thus, this cluster was labeled as the “lowest level is dominant” theme for the number of project content words feature. The goal of step 2 was to deriving user patterns through the observed dominant themes. The 70% rule was reached based on several trials of experiments. Setting a higher percentage bar left fewer dominant themes for us to make inferences, while a lower percentage bar was too lenient and hardly produced distinctive traces for each cluster. Thus, we settled on 70%.

It is worth noting that although we had one 2-level feature and eight 3-level features, which in theory should produce 42 features in total, only 30 dominant themes emerged from this study. If a feature under a certain setting did not display a dominant theme, it was dropped from that particular cluster.

Lastly, the dominant themes for each cluster were combined together to represent a usage pattern. Again taking the
2nd cluster when \( k = 3 \) as an example, its final usage pattern was: \{the number of projects = more than one AND the number of words in project overview = none or a few AND the number of words in project content = none or a few AND the number of resources in project = none or a few AND the number of student visits = a few or many AND the number of projects being copied = none\}.

**Step 3. Mining of frequent user patterns.** Frequent itemsets mining (Han & Kamber, 2006) was used to find the user patterns that most often occurred together, in particular identifying the itemsets that exist in more than a certain proportion of the entire dataset. In data mining language, this proportion threshold is called support. In this study, we set the minimum support at 10%. This means that in order to be considered as a frequent user pattern, a combination of feature themes needed to appear six times or more in the 59 usage patterns generated in Step 2.

An Open Source data mining tool, Weka, was then used for frequent itemsets mining, and identified 24 1-itemsets, 110 2-itemsets, 190 3-itemsets, 182 4-itemsets, 102 5-itemsets, 31 6-itemsets, and four 7-itemsets frequent user patterns. For example \{number of projects = one AND number of words in project overview = high AND number of words in project content = high AND number of project resources = high AND number of student visits = zero\} is one of the discovered 5-item frequent user patterns.

**Step 4. Selecting final user clusters.** The final user clusters were selected among the frequent itemsets. Selecting meaningful and useful patterns from the large number of frequent itemsets can be a difficult and subjective process. In this study, four principles were used to guide the selection process:

1. **Mutual exclusiveness.** The selected frequent itemsets should not overlap in any of its individual feature’s theme. This guaranteed that the final user clusters had no conflicting patterns and thus any user would belong to only one final cluster.

2. **Balance.** Balanced cluster size (\( N \)) was preferred; a cluster that was too small (\( N < 100 \)) or too large (\( N > 200 \)) was not selected even if it met all the other principles.

3. **Comprehensiveness.** Recall that the user feature space allowed for three roles: project authoring, resource usage, and navigation. Ideally, the final selected frequent itemsets should exhibit distinctive themes in all three aspects of the feature space. If it cannot be met, a frequent itemset covering more roles was preferred.

4. **Maximum.** Given two similar user patterns that both meet the other three principles, the pattern containing more items (pairs of features and themes) in it was preferred. If the two patterns contained the same number of items, then the one with more users in it was preferred.

The four clustering and cluster pruning steps produced three user clusters, as shown in Table 2. Each user cluster represented a distinctive user pattern and the defining indicators are noted with asterisks. Those indicators are the dominant themes of each cluster. As part of the data post-processing phase, the next section provides an interpretation of and labels for the three clusters, based on their overarching characteristics.

**Phase 3 -- Data post-processing 1: Interpreting the clustering result**

**Cluster 1: Key brokers (\( N = 108 \)).** Teachers in this group were frequent browsers, had verbose projects, and created projects that attracted visits from other people. Of all three groups, this group scored relatively high on other measures, except for the maximum number of student projects, which was lower than cluster 2. This group did not necessarily share every single project with the public, but was careful in selecting what to share, suggesting that teachers in this group gave serious thought to their IA projects. If the IA is viewed as a learning community, teachers in Cluster 1 were the stickiest and key brokers because they appeared to be willing to observe and learn from others and also give back to the community.

**Cluster 2: Insular classroom practitioners (\( N=114 \)).** This group of teachers did not create high-quality projects, as they were characterized by few resource links, limited overview, and little content. Meanwhile they did not visit the IA or browse others IA projects as often as teachers in cluster 1, nor did they copy other teachers’ projects for their own use. In spite of the lack of enthusiasm for creating IA projects, they appeared to implement their IA projects in classroom teaching. Students viewed their projects at least once; 50% of the teachers in this group had projects viewed by the students five times or more, and in addition, 30% had projects viewed by the students 10 times or more. Given their behaviors, this group is dubbed the insular classroom practitioners.
Cluster 3: Inactive islanders (N=126). This group of teachers only published one IA project each. These published projects were apparently good when judged by three project authoring measures: a medium amount of text in the overview, relatively verbose project body, and a reasonable number of resource links. In terms of navigation, this group appeared to be relatively inactive, as it was low in all three navigation measures. We speculate that the fact that users did not explore the IA as much as those in cluster 1 may have affected their knowledge of using the IA as well as their skills in creating quality IA projects. The IA was designed to allow teachers to collect and reuse web resources, and borrow curricular ideas from each other. Since this group was isolated from others and showed little navigation, it was dubbed inactive islanders.

### Table 2. Final User Clusters

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Overall</th>
<th>Cluster 1 (N=108)</th>
<th>Cluster 2 (N=114)</th>
<th>Cluster 3 (N=126)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role</td>
<td>name</td>
<td>range</td>
<td>mean</td>
<td>median</td>
</tr>
<tr>
<td>Project authoring</td>
<td>Number of projects</td>
<td>1~9</td>
<td>2.34</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Project overview words</td>
<td>0~293</td>
<td>20.57</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Project content words</td>
<td>0~2843</td>
<td>153.60</td>
<td>67.50</td>
</tr>
<tr>
<td>Resource links</td>
<td>0~286</td>
<td>4.24</td>
<td>2.50</td>
<td>0~26</td>
</tr>
<tr>
<td>Project usage</td>
<td>Max student visits</td>
<td>0~1022</td>
<td>14.41</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Max peer visits</td>
<td>0~160</td>
<td>2.37</td>
<td>0</td>
</tr>
<tr>
<td>Navigation</td>
<td>Visits to the IA</td>
<td>1~57</td>
<td>8.18</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Project browses</td>
<td>0~88</td>
<td>10.51</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Copied projects</td>
<td>0~18</td>
<td>0.49</td>
<td>0</td>
</tr>
</tbody>
</table>

**Note.** The indicators with asterisks are the defining features for that cluster.

Phase 3 -- Data post-processing II: Triangulating the clustering results

This section describes our second data post-processing efforts, in which we validated cluster interpretations with an additional triangulation study.

When teachers first register for their free IA account, they are asked to optionally answer two user profile questions: years of teaching experience (0 ~ 3, 4+), and comfort level with technology (on a scale of 0 “low” to 4 “high”). For the three clusters of teachers (N=348), 116 reported their years of teaching, and 292 reported their comfort level with technology. Tables 3 and 4 show how these profile items are distributed in the three user clusters.

The tables show that the key brokers cluster had a larger proportion of tech-savvy teachers than the other two groups, and the insular classroom practitioners group mostly consisted of novice teachers. To test whether this is a random effect, a chi-square test and an exact test were used as preliminary analyses to evaluate the frequency distributions of the demographic profile across the different clusters. The chi-square test is used when the sample size is large, while the exact test is used when any cell has small (< 5) or 0 counts.

### Table 3. Teacher clusters by teaching experience

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Novice (1 ~ 3 years)</th>
<th>Experienced (≥ 4 years)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key brokers</td>
<td>24 (40%)</td>
<td>36 (60%)</td>
<td>60</td>
</tr>
<tr>
<td>Insular classroom practitioners</td>
<td>16 (84%)</td>
<td>3 (16%)</td>
<td>19</td>
</tr>
</tbody>
</table>
Table 4. Teacher clusters by comfort level with technology

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Low (0 - 1)</th>
<th>Medium (2)</th>
<th>High (3 - 4)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Key brokers</em></td>
<td>17 (18%)</td>
<td>36 (37%)</td>
<td>44 (45%)</td>
<td>97</td>
</tr>
<tr>
<td><em>Insular classroom practitioners</em></td>
<td>16 (18%)</td>
<td>51 (58%)</td>
<td>21 (24%)</td>
<td>88</td>
</tr>
<tr>
<td><em>Inactive islanders</em></td>
<td>17 (16%)</td>
<td>51 (48%)</td>
<td>39 (36%)</td>
<td>107</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>138</strong></td>
<td><strong>104</strong></td>
<td><strong>292</strong></td>
</tr>
</tbody>
</table>

An exact test showed that the probability distribution of the teaching experience was significantly different among the three groups ($p < .01$). A chi-square test showed that the probability distribution of the comfort level with technology was also significantly different among the three groups ($\text{chi-square}=10.42, p < .05$).

Given these results, we fitted a multinomial logistic regression model to further explore how teachers’ teaching experience and technology comfort level were related to their online behaviors, with teachers’ cluster labels set as the response variable, and their profile data as the explanatory variable. In Table 5, $B$ is the estimated coefficient relative to the reference cluster (*key brokers*), while $\exp(B)$ is the exponentiation of $B$, or the odds ratio of being in this group relative to the reference cluster, in this case, the *key brokers*. Finally, the percentage is calculated from $\exp(B)$, and indicates the change of predicted odds in percentages as compared to the reference group. Positive numbers are increases, while negative numbers are decreases.

Table 5. Multinomial logistic regression analysis of the impact of teaching experience and comfort level with technology on users’ online behaviors

<table>
<thead>
<tr>
<th>Teaching Experience</th>
<th>B</th>
<th>p-value</th>
<th>Exp(B)</th>
<th>percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Insular classroom practitioners</em></td>
<td>-2.08</td>
<td>.00</td>
<td>.13</td>
<td>-87%</td>
</tr>
<tr>
<td><em>Inactive islanders</em></td>
<td>-1.02</td>
<td>.02</td>
<td>.36</td>
<td>-64%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comfort Level with Technology</th>
<th>B</th>
<th>p-value</th>
<th>Exp(B)</th>
<th>percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Insular classroom practitioners</em></td>
<td>-.45</td>
<td>.03</td>
<td>.64</td>
<td>-36%</td>
</tr>
<tr>
<td><em>Inactive islanders</em></td>
<td>-.15</td>
<td>.45</td>
<td>.86</td>
<td>-14%</td>
</tr>
</tbody>
</table>

Notes. The *key brokers* cluster is the reference (baseline) cluster.

As shown in Table 5, for teaching experience, the coefficient for the *insular classroom practitioners* relative to *key brokers* is -2.08, in other words, the predicted odds of being categorized as an *insular classroom practitioner* rather than a *key broker* would decrease by 87% ($p < .01$). The coefficient for *inactive islanders* relative to *key brokers* is -1.02, thus the predicted odds of being in the *inactive islanders* cluster rather than the *key brokers* cluster would decrease by 64% ($p < .05$). In sum, an experienced teacher is expected to have a higher chance being in the *key brokers* cluster than in the other two types.

As shown in Table 5, for comfort level with technology, for every one unit increase (from low to medium, or from medium to high), the coefficient for being an *insular classroom practitioners* relative to *key brokers* is -.45, thus the predicted odds of being in the *insular classroom practitioners* cluster rather than the *key brokers* cluster would decrease by 36% ($p < .05$). For every one unit increase in technology comfort level, the coefficient of being in the *inactive islanders* cluster relative to the *key brokers* cluster would be expected to -.15, and the predicted odds of being in the *inactive islanders* cluster rather than the *key brokers* cluster would decrease by 14%, but this difference failed to achieve statistical significance.

In sum, the multinomial logistic regression showed strong relationships between teachers’ characteristics and their online behaviors as described by user clusters. Specifically, teachers with more teaching experience were more likely to be *key brokers*, and those with less teaching experience were more likely to be *inactive islanders*. Teachers who
were more comfortable with technology were more likely to be *key brokers* and were least likely to be *insular classroom practitioners*.

**Conclusions, limitations, and future work**

This research examined and analyzed teachers’ online behaviors in the context of a digital library tool, the Instructional Architect. First, an educational data mining approach, clustering, was applied to identify different groups of IA teacher users according to their diverse online behaviors. A user model consisting of nine features was identified and fed into a LCA model, clustering IA teacher users into three groups, labeled *key brokers, insular classroom practitioners, and inactive islanders.*

Second, a triangulation study examined relationships between teachers’ profile data and their usage patterns. This analysis showed strong relationships between teachers’ characteristics and their online behaviors as described by user clusters. Specifically, teachers with more teaching experience were more likely to be *key brokers,* and those with less teaching experience were more likely to demonstrate ineffective use of the IA. Teachers who were more comfortable with technology were more likely to be *key brokers* and were least likely to be *insular classroom practitioners.* Such results show that effective usage of the Instructional Architect requires both pedagogical knowledge (gained through experience teaching) and technological knowledge. This finding helps to predict which kinds of teachers are more likely to adapt technology tools such as digital libraries, and more importantly, how to help teachers become more effective digital libraries users.

Three areas are proposed for future work. First, although LCA is alleged to outperform K-means, no competing clustering algorithm has been implemented to justify this choice. Secondly, previous work showed that greater use of the IA occurs in geographical areas where teacher professional development workshops using the IA have been conducted (Khoo et al., 2008; Xu, Recker, & Hsi, 2010). This suggests that workshop participants have a higher chance of becoming *sticky* users. Therefore, teachers who participated in such workshop can be singled out for detailed analysis, as their distribution among clusters is predicted to be different. Finally, the third stage of KDD, evaluation and interpretation, could be conducted in a more comprehensive fashion. For example, the survey information filled out by workshop participants could be used to triangulate the clustering results, providing evidence for why and how the teachers like and dislike the IA.

Despite the current challenges, the field of educational data mining is making progress towards standardizing its procedures for tackling educational problems. This research shows that teachers’ use of online resources can be studied by productively using web usage data and employing data mining approaches to investigate digital library problems in innovative ways.

**Acknowledgements**

This material is based upon work supported by the National Science Foundation under grant #0840745. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the teacher users of the IA.

**References**


Analyzing Interactions by an IIS-Map-Based Method in Face-to-Face Collaborative Learning: An Empirical Study

Lanqin Zheng, Kaicheng Yang* and Ronghuai Huang
School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing, 100875, China
//bnuzhenglq@bnu.edu.cn//yangkc_beijing@foxmail.com//huangrh@bnu.edu.cn

*Corresponding author

ABSTRACT
This study proposes a new method named the IIS-map-based method for analyzing interactions in face-to-face collaborative learning settings. This analysis method is conducted in three steps: firstly, drawing an initial IIS-map according to collaborative tasks; secondly, coding and segmenting information flows into information items of IIS; thirdly, computing attributes of information flows and analyzing relationships between attributes and group performance. An example illustrates how the methodology uncovers the interaction process based on information flows. The empirical study aims to validate the effectiveness of this method through thirty groups’ interactions. The result indicates that quantity of activation of the targeting knowledge network can predict group performance and the IIS-map-based analysis method can analyze interactions effectively. The primary contribution of this paper is the methodology for analysis of interactions based on information flows.

Keywords
Collaborative learning, IIS-map-based analysis method, Interaction analysis, Information flows, Knowledge construction

Introduction
In the past decade, more and more attention has been paid to collaborative learning. A major theme in the collaborative learning field is why some groups are more successful than others (Barron, 2003; Suthers, 2006). Lately, researchers have sought to address this issue by analyzing interaction processes in collaborative learning reasoning that human cognition is based on interactions between individuals and social context or community (Engeström, 1987). Various methods have been developed in previous research to analyze interactions. The following analytic methods have been widely used: (a) Conversation analysis (Sacks, 1962,1995), identifying closings and openings of action sequences (Zemel, Xhafa, & Stahl, 2005); (b) Social network analysis (Wasserman & Faust, 1994), investigating patterns of interaction (de Laat, Lally, Lipponen, & Simons, 2007) and examining the response relations among participants during online discussions (Aviv, Erlich, Ravid, & Geva, 2003; De Liddo et al., 2011); (c) Content analysis (Chi, 1997), using coding schemes to categorize and count user actions to analyze argumentative knowledge construction (Weinberger & Fischer, 2006), evidence use for the knowledge building principles (van Aalst & Chan, 2007), depth of understanding (Zhang, Scardamalia, Reeve, & Richard, 2009); (d) Sequential analysis, using transitional state diagrams to compute transitional probabilities between coded discourse moves in argumentation (Jeong et al., 2011). Each method has limitations. Table 1 summarizes the analysis approaches, focus and limitations of the different methods.

Table 1. Comparison of different analysis methods

<table>
<thead>
<tr>
<th>Analysis methods</th>
<th>Analysis approaches</th>
<th>Focus</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversation analysis</td>
<td>Qualitative</td>
<td>Turn-taking</td>
<td>Without conveying the dynamics of conversation</td>
</tr>
<tr>
<td>Social network analysis</td>
<td>Quantitative</td>
<td>Relationships of members</td>
<td>Irrespective of social background</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The conclusions are generally personalized.</td>
</tr>
<tr>
<td>Content analysis</td>
<td>Qualitative, quantitative</td>
<td>Speech acts</td>
<td>Coding is based on subjective judgments.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Neglect of domain knowledge.</td>
</tr>
<tr>
<td>Sequential analysis</td>
<td>Quantitative</td>
<td>Discourse moves</td>
<td>Ignoring knowledge construction</td>
</tr>
</tbody>
</table>

At present, the most often used method is content analysis (Strijbos & Stahl, 2007). Content analysis technique is defined as “a research methodology that builds on procedures to make valid inferences from text” (Rourke, Anderson, Garrison, & Archer, 2001). The essential step of content analysis is to code discussions according to the selected coding scheme. However, different researchers put forward different coding schemes. Well-known examples include content coding schemes for the analysis of the learning process in computer conferencing (Henri,
1992), co-construction of understanding and knowledge (Zhu, 1996), the social construction of knowledge in computer conferencing (Gunawardena, 1997), the social presence in the community of inquiry (Rourke, 1999), the collaborative construction of knowledge (Veerman & Veldhuis-Diermanse, 2001; Pena-Shaff & Nicholls, 2004), the cognitive presence in community of inquiry (Garrison et al., 2001), the teaching of the community of inquiry (Anderson et al., 2001), and argumentative knowledge construction (Weinberger & Fischer, 2006). De Wever et al. (2006) compared 15 content analysis instruments from the perspective of the theoretical base, unit of analysis and inter-rater reliability and pointed out that existing analysis instruments need to be improved.

Every content analysis scheme uses its own specific unit of analysis and data type. The analysis units are not identical in a variety of coding schemes, such as messages (Gunawardena et al., 1997), sentences (Fahy et al., 2001), paragraphs (Hara et al., 2000) and thematic units (Henri, 1992). The selection of the unit of analysis is very challenging for researchers. Although many researchers use “thematic unit” as the unit, the categorization standard of the “thematic unit” is very ambiguous. The complexity of interaction makes researchers use different vocabularies to code transcripts into different speech acts. For example, Fahy et al. (2001) coded transcripts into five kinds of speech acts (question, state, reflection, comment, and quote). The coding scheme developed by Pena-Shaff and Nicholls (2004) consisted of eleven kinds of speech acts (question, reply, clarification, interpretation, conflict, assertion, consensus building, judgment, reflection, support and other). Pilkington (2001) believes that coding schemes may categorize at too coarse a level to distinguish real communicative differences, or they may be too fine-grained to represent similarities. Porayska-Pomsta (2000) argues that categorizing speech acts is not useful in modeling teacher’s language and cannot account for the phenomena encounter in the dialogues. Furthermore, coding assigns each speech act an isolated meaning and does not record the indexicality of the meaning or contextual evidence (Suthers et al., 2010).

In addition, the difficulty with content analyses of communications stems from a lack of guidelines for performing them validly and reliably (Rourke et al., 2001; Strijbos et al., 2006). Rourke et al. (2001) also discussed the importance of inter-rater reliability in the method of content analysis and pointed out that many researchers did not report coder reliability. Strijbos et al. (2006) believed that researchers should be cautious of the statistical test results when they did not report reliability parameters. The works of Dillenbourg (1999) and Stahl, Koschmann, and Suthers (2006) call for the need to develop process-oriented methodologies to analyze interactions.

We believe that coding interaction transcripts into speech acts is very difficult because purposes of human’s speech acts are implicit; thus the identification of speech acts is very subjective. Simply focusing on the explicit speech acts will lead to ignorance of an individual’s knowledge construction. This study proposes an innovative method to analyze interactions in face-to-face collaborative learning. This method is IIS-map-based analysis method because it uses the IIS map. The whole study aims to validate the IIS-map-based analysis method that is used to analyze interactions and predict group performance. The empirical study is conducted to explore the effectiveness of the IIS-map-based analysis method and to verify hypotheses.

Methodology: IIS-map-based analysis method

Modeling and representing the collaborative learning system by IIS-map-based analysis method

You (1993) believes that the instructional system is a complex non-linear system that assumes cause and effect are associated disproportionately and the whole is not simply the sum of the properties of its parts. In addition, complex systems have an “emergence” property. Emergent properties arise at a particular level of system description by virtue of the interaction of relatively simple lower-level components – but cannot be explained at this lower level (Damper, 2000). Kapur et al. (2011) believes that the group is a complex system and convergence in group discussions is an emergent behavior arising from interactions between group members. Therefore the instructional system and collaborative learning system are both complex systems with characteristics of non-linearity and emergence. The complex systems cannot be understood by only analyzing visible factors such as teaching methods, various kinds of media, etc. To deeply understand various complex pedagogy phenomena and their effects, researchers should focus on the information flow within the system and its characteristics, as well as relationships between information flows and functions of the system (Yang & Zhang, 2009). We argue that the instructional system is an abstract information system. The collaborative learning system is a subsystem of instructional system, so it is also an information system. The function of a collaborative learning system is collaborative construction of knowledge by group members.
Information processing and knowledge construction are closely intertwined in the learning process (Wang et al., 2011). The cognitive processes involved in knowledge construction are selecting relevant information from what is presented, organizing selected information into a coherent representation, and integrating presented information with existing knowledge (Mayer, 1996). The interconnection of the prior knowledge with the new information can result in reorganization of the cognitive structure, which creates meaning and constructs knowledge. Learning is a generative process of constructing meaning by linking existing knowledge and incoming information (Osborne & Wittrock, 1983). Based on the theoretical foundations, we argue that the nature of knowledge construction is to encode and decode information implicitly. Therefore information makes significant contributions to knowledge construction. Accordingly, the analytic focus is identified as information flows of the collaborative learning system. The information flow is defined as the output information of group members in the interaction process. The information flows between private information owned by each individual and the information shared by group members.

In order to represent and analyze the collaborative learning system, a concept model is designed (see Figure 1). In this concept model IPL denotes the information processing of learners. IPL₁, IPL₂, IPL₃, IPL₄ denote information processing of multiple learners in one group. The internal information processes of IPL are not directly observable. However, the input and output information of IPL are visible. So {X} denotes the input information of IPL and {Y} denotes the output information of IPL. Because the output information {Y} is used for the purpose of sharing information, {Y} is abstractly generalized into an information set. This abstract information set is defined as Interactional Information Set (IIS). IIS is for sharing information in the interaction process. Thus {Y} is regarded as the input information of IIS and {X} is regarded as the output information of IIS. Vygotsky (1978) argue that learning takes place inter-subjectively through social interaction before it takes place intra-subjectively. IIS is generated and formed when information are externalized and shared in the social interaction process. Therefore IIS can account for social aspects of learning. We argue that IIS can represent the outcome of internal information processing of IPL. Because knowledge is constructed through processing information implicitly, some characteristics of IIS are closely related to the quality of co-construction of knowledge. The whole collaborative learning system is a functional coupling system which consists of IPL, {X}, {Y} and IIS.

![Figure 1. The concept model of the collaborative learning system](image)

**Coding and representing the input information of IIS**

According to the concept model, three kinds of objects need to be represented. The first kind of object is the input information of IIS, namely {Y}. Because {X} is the input information of IPL and {Y} is the output information of IPL, {X} can be finally embodied and represented by {Y}, the analysis of {X} is unnecessary and the analysis focus is {Y}. In order to analyze the collaborative learning system, the attributes of input information of IIS need to be defined. These attributes include time, information processing of learners (IPL), cognitive levels, information types,
representation formats, knowledge network sub-map, annotation and the quality of information. Table 2 below shows the definition of each attribute. The coding format of input information items of IIS is defined as: 

\[
\text{<time><IPL_i><cognitive level><information type><representation format><knowledge network sub-map> [annotation] [the quality of information].}
\]

The symbol “< >” denotes the required item and “[ ]” denotes the optional item. Information flows in the interaction process will be represented and transformed into this kind of format.

**Table 2. The definition of each attribute**

<table>
<thead>
<tr>
<th>Attributes of input information of IIS</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Time means the start time of each information flow.</td>
</tr>
<tr>
<td>IPL_i</td>
<td>“i” in the IPL_i is used to distinguish different learners, such as IPL_1 means the information processing of the first learner, IPL_2 means the information processing of the second learner, and so on.</td>
</tr>
<tr>
<td>Cognitive levels</td>
<td>Cognitive levels of IPL_i include discriminating, recalling, understanding and applying.</td>
</tr>
<tr>
<td>Information types</td>
<td>Information types include description of objectives, context, knowledge semantics, answers and questions, facts and examples, management instructions, related information, unrelated information.</td>
</tr>
<tr>
<td>Representation formats</td>
<td>The values of representation formats include text (T), sound(S), graph (G), photo (P), table (Tb), video (V), animation (A), object (O), body language (B).</td>
</tr>
<tr>
<td>Knowledge network sub-map</td>
<td>The knowledge network sub-map is represented by knowledge and their relationships.</td>
</tr>
<tr>
<td>Annotation</td>
<td>The annotation is to introduce the main idea of each information flow.</td>
</tr>
<tr>
<td>The quality of information</td>
<td>The quality of information can be differentiated good, moderate, inferior. The defaulted value is good.</td>
</tr>
</tbody>
</table>

**Visualizing information flows into sequences of IIS**

In order to clearly represent information flows in the interaction process, the input information items of IIS are visualized into sequences according to the coding format. The sequences of input information items are the second kind of object that needs to be represented. Collaborators’ information items are chronologically on the timeline or under the timeline. The sequences of information items in face-to-face collaboration are linear sequences. Figure 2 shows the portion of sequential representation of IIS. The four long strands specify the portion of information sequences being shared at different time.

![Figure 2. The portion of sequential representation of IIS](image)

**The representation of IIS —— Knowledge Network Maps**

The third kind of object need to be represented is IIS. We believe that IIS is closely related to the outcome of collaborative construction of knowledge. However, IIS is an abstract information set and it is not the real storage medium. Thus in order to represent IIS, a knowledge network map is used for visualizing knowledge and their relationships. From the beginning the knowledge network map is empty. As the interaction proceeds, the knowledge
network map is gradually enlarged. Finally, it becomes the stable knowledge network map. The IIS-map is the knowledge network map with marks. The marks refer to attributes of information flows. The analysis of interactions can be reflected by the IIS-map. So this method is known as the IIS-map-based method. IIS-maps serve as abstract transcripts that can represent and visualize the interaction process.

The Steps of IIS-map-based analysis method: An example

This section illustrates the steps of the IIS-map-based analysis method by using the example of co-constructing knowledge of graphs in the data structures curriculum. The IIS-map-based analysis method is conducted in three steps:

Firstly, draw an initial IIS-map according to collaborative learning tasks. In this example, collaborative learning tasks are to design the algorithm of touring the campus so as to provide services for visitors. The learning objective of this task is to acquire storage structures of a graph, the algorithm of the minimum spanning tree and the shortest path of a graph. The detailed descriptions of tasks are as follows:

1. Design a campus plan, including the south gate, Jingshi Square, gymnasium, playground, library, technology building, and art building. How many kinds of storage structures can represent this problem? Explain the relative merits respectively.
2. Provide some information such as the name, code number and introduction of each scenery spots.
3. If sewer pipes are laid between each scenic spot, please design how to lay them down at the minimum cost.
4. If a visitor wants to tour from the Jingshi Square to the art building, please design the shortest path between them; In addition, please design the second shortest path between these two scenic spots.
5. If a visitor wants to tour the library and the technology building, please design the shortest path between the Jingshi Square and the art building, passing by the library and the technology building.
6. If a visitor sets out from the south gate and tours each scenic spot only once, and then leaves via the south gate, what will be the shortest tour path?

The initial IIS-map can visually represent the domain knowledge structure. Generally speaking, the initial IIS-map represents teachers’ understanding of domain knowledge. The initial IIS-map is drawn according to the knowledge modeling norm (Yang, 2010a). This norm specifies seven categories of knowledge and eighteen kinds of relationships of knowledge. For more information refers to (Yang, 2010a). Figure 3 shows the portion of the initial IIS-map, in which the node represents the knowledge and the edge represents mutual relationship of the knowledge. Of course any kind of accepted norm can be adopted to draw the initial map.
Secondly, code and segment information flows into information items sequences of IIS according to the representation format. Specifically, the segmentation is in chronological order and it is based on the change of attributes of information items. The rules for segmentation and coding information flows are as follows:

1. When contributors of information change, the information flow will be segmented.
2. When cognitive levels change, the information flow will be segmented. The definition of cognitive level originates from the theory of instructional objective classification (Yang, 2010a). Discriminating, recalling and understanding are the same level. When the level changes into “applying” the information flow will be segmented because “applying” means a higher cognitive level.
3. When types of information change, the information flow will be segmented.
4. When the knowledge network sub-map changes, the information flow will be segmented.
5. The values of representation format do not influence the segmentation of information flows because there are many representation formats in each information flow such as text, sound, graphics and body language.

All information flows in the interaction process need to be coded and segmented independently by two raters according to the rules. Table 3 shows fragments of information flows that originate from face-to-face interactions of a group. We code and segment information flows of Table 3 into sequences of information items of IIS, as shown in Figure 4.

### Table 3. Fragments of information flows

<table>
<thead>
<tr>
<th>Time</th>
<th>IPL</th>
<th>Information flows</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>IPL1</td>
<td>“I think that the traveling salesman problem can use the algorithm of enumeration.”</td>
</tr>
<tr>
<td>57</td>
<td>IPL2</td>
<td>“Let me think about it. The traveling salesman problem can use the retrospective algorithm.”</td>
</tr>
<tr>
<td>1:04</td>
<td>IPL3</td>
<td>“I believe that the traveling salesman problem can use the algorithm of dynamic programming.”</td>
</tr>
<tr>
<td>1:12</td>
<td>IPL2</td>
<td>“Actually the traveling salesman problem is NP-complete.”</td>
</tr>
<tr>
<td>1:16</td>
<td>IPL1</td>
<td>“Branch and bound algorithm can also solve the traveling salesman problem.”</td>
</tr>
<tr>
<td>1:19</td>
<td>IPL2</td>
<td>“Yes, you are right. Great!”</td>
</tr>
<tr>
<td>1:25</td>
<td>IPL3</td>
<td>“A* algorithm is also a good solution.”</td>
</tr>
<tr>
<td>1:31</td>
<td>IPL1</td>
<td>“Is there any other algorithm?”</td>
</tr>
<tr>
<td>1:39</td>
<td>IPL3</td>
<td>“There is a greedy algorithm for solving the traveling salesman problem.”</td>
</tr>
<tr>
<td>1:43</td>
<td>IPL2</td>
<td>“What’s the greedy algorithm?”</td>
</tr>
</tbody>
</table>

![Figure 4. Portion of the sequences of input information items of IIS](image-url)
Thirdly, compute attributes of information flows and generate the knowledge network map with marks. The marks refer to the quantity of activation. The quantity of activation is the information entropy that is generated by activating. We believe that some of the knowledge network maps are activated when information flows map the IIS. As long as the information flow has the knowledge network sub-map attribute, the knowledge is activated. The quantity of activation is an abstract attribute of information flows. It is designed to represent the quality of knowledge construction. Yang (2010b) proved that the quantity of activation of knowledge was positively correlated with learning outcomes in the context of classroom teaching. The higher the quantity of activation of knowledge is, the better learning outcomes will be. The algorithm of the quantity of activation of knowledge was designed by Yang (2010b) as formula (1):

$$A_i = \sum F \cdot \frac{\log(d + 2) \cdot r}{\log(n \cdot (D - d + 2))}$$

(1)

Where:
- $F$ is an adjustable parameter. $F = 1$ when there is $\leq 7$ information items between the current information item and the preceding information item that activated the current knowledge. $F = \frac{k + 1}{(2k + 1)}$ when there are more than seven information items. The value of $k$ is equal to the number of information items between the current information item and the preceding information item that activated the current knowledge subtracting seven.
- $\log(d + 2)$ denotes the activation entropy and “$d$” is the number of the activated edges.
- $\log(n \cdot (D - d + 2))$ denotes the complexity. “$n$” denotes the categories of edges that are not activated. $D$ denotes the total number of edges that connected with the vertex.
- $r$ is an adjustable parameter. $r = 1$ when the vertex is directly activated. $r = \frac{1}{d}$ when the vertex is activated only once by its adjacent vertex. $r = \frac{1}{d^2}$ when the vertex is activated twice by its adjacent vertex.

The IIS-map-based analysis method aims to model and represent the collaborative learning system by information flows. It can visualize the interaction process and represent information flows onto the IIS map. The IIS-map-based analysis method regards the knowledge network map as the objective reliance body. The main content loaded by information flows is reflected and represented by the knowledge network map. This reduces the arbitrariness of coding discourse to a large extent, which makes the segmentation of information flows more objective and scientific.

An Empirical study

This empirical study aims to verify if IIS-map-based analysis method can analyze interactions and some attributes of information flows can predict the group performance. The group performance is measured as the quality of knowledge construction, which can be calculated according to formula (2):

$$X = \frac{P \times (\sum X_{\text{posttest}} - \sum X_{\text{pretest}})}{N \times \sqrt{CV}}$$

(2)

Where $P$ denotes the difficulty coefficient of test items. The difficulty coefficient equals the ratio of the average score to the full mark of test items. $CV$ denotes the coefficient of variation. It equals the percentage of the ratio of standard deviation to mean of score. $X_{\text{posttest}}$ and $X_{\text{pretest}}$ respectively represent the score of pretest and posttest. $N$ denotes the number of group members. This algorithm of group performance will be more objective in contrast to only computing the mean deviation between the pretest and posttest.

The definitions of attributes of information flows

Interactions are very complex and some groups may deviate from learning objectives in the interaction process. Therefore on the IIS-map there is the targeting knowledge network map which is composed of the targeting knowledge. The targeting knowledge network can be selected and identified by teachers according to collaborative learning objectives. In this study we focus on one of the important attributes of information flows, namely characteristics of the targeting knowledge network map. Several indicators are designed to compute the
characteristics of targeting knowledge network map, including the quantity of activation, average degree, average path length, density and degree entropy of the targeting knowledge network map.

**Quantity of activation of the targeting knowledge network map**

Collaborative learning is different from the instruction which emphasizes the delivering of new knowledge to students. The precondition of collaborative learning is that the participants understand each other enough to accomplish their work (Stahl, 2011b). Furthermore, collaborative learning emphasizes collaborative construction of meaning among group members. Therefore the quantity of activation of the targeting knowledge network map is used for indicating the quality of knowledge co-construction. Many algorithms are tested. Finally, quantity of activation of the targeting knowledge network map is defined as the sum of quantity of activation of the targeting knowledge on the IIS map, as is shown in formula (3).

\[ A = \sum_{i=1}^{N} A_i \]  

(3)

Where \( A_i \) denotes the quantity of activation of the targeting knowledge which can be calculated according to formula (1). \( N \) denotes the number of vertices of the targeting knowledge network map.

**Average degree**

The average degree indicates the mean number of neighbors per vertex of the targeting knowledge network. The average degree can be calculated using formula (4):

\[ d = \frac{2 \times E}{N} \]  

(4)

Where \( E \) denotes the total number of edges. \( N \) denotes the total number of vertices.

**Average path length**

The average path length indicates the scale of the targeting knowledge network. The average path length can be calculated using formula (5):

\[ L = \frac{2}{N \times (N - 1)} \sum_{i,j} d_{ij} \]  

(5)

Where \( N \) denotes the total number of vertices. \( d_{ij} \) denotes the quantity of edges of the shortest path between any two vertices.

**Density**

Density can indicate the interconnectivity and closeness of associations of the targeting knowledge. The density of the vertex is calculated as the proportion of the number of actual edges to the number of possible edges. The density can be calculated using formula (6):

\[ D = \frac{1}{N} \sum_{i=1}^{N} \frac{2 \times E_i}{k_i \times (k_i - 1)} \]  

(6)

Where \( E_i \) denotes the number of actual edges. \( k_i \) denotes the degree of the vertex. \( N \) denotes the total number of vertices.
**Degree entropy**

The degree entropy shows the heterogeneity of targeting knowledge network. We adopt the algorithm of degree entropy in the complex network (Ferrer & Sole, 2003). It is defined as formula (7):

\[ H = - \sum_{k=1}^{\infty} p_k \log p_k \]  

(7)

Where \( p_k \) is the frequency of vertices having degree “\( k \)” and \( \sum_{k=1}^{\infty} p_k = 1 \).

We can apply these measures to provide insight into the relationship between attributes of information flows and group performance.

**Hypotheses**

We suppose that the following attributes of information flows can predict group performance. So five hypotheses of this study are proposed:

H1: The quantity of activation of the targeting knowledge network map can predict group performance.

H2: The average degree of the targeting knowledge network can predict group performance.

H3: The average path length of the targeting knowledge network can predict group performance.

H4: The density of the targeting knowledge network can predict group performance.

H5: The degree entropy of the targeting knowledge network can predict group performance.

As long as one hypothesis is verified, the effectiveness of the IIS-map-based analysis method is validated.

**Experiment Procedure**

The experiment used pretest posttest research design. The section presents the step-by-step process for this study:

1. Setting collaborative learning objectives and selecting some knowledge as learning objects. In this study, tasks were originated from data structures. The detailed descriptions were shown in the section of the first step of the IIS-map-based analysis method. Teachers designed the collaborative task according to the learning objectives. This task was based on a real-life authentic situation, which was not explicitly taught in the course of data structures. Meanwhile, subjective testing items of the pretest and posttest were designed according to learning objective, which could be used to evaluate group performance. The detailed descriptions of pretest and posttest are in Appendix. The targeting domain knowledge included storage structures of a graph, the algorithm of the minimum spanning tree and the shortest path of a graph.

2. Drawing an initial IIS-map of the targeting domain knowledge. The initial IIS-map was composed of targeting knowledge and their relationships. The initial IIS-map was viewed as the standard map, and it represented what teachers expected to discuss in the interaction process (Figure 3).

3. Recruiting subjects and dividing them into different groups. The experiment selected thirty groups of subjects. There were three or four students in each group. The group members were divided into each group randomly. Then the subjects were placed in different laboratory rooms and collaborated via face to face. Group members collaborated for about two hours to solve the problem and designed at least one algorithm. Each group wrote down procedures of the algorithm in the last collaborative phase as the artifact. The pretest preceded the collaboration, followed immediately by the posttest. In order to record the authentic interaction process, we videoed the entire interaction process of thirty groups, and each group formed the total of approximately 120 minutes of video file in this experiment, according to which the initial knowledge network map was modified. In the experiment, the tasks were the same for the thirty groups, as were the questions for pretest and posttest.

4. Coding, segmenting and transforming information flows into sequences of information items according to the segmentation rules, as shown in Figure 4.

5. Computing the attributes of information flows, such as the quantity of activation, average degree, average path length, density and degree entropy of the targeting knowledge network, then generating the knowledge network map with marks, as shown in Figure 5. The numbers beside the knowledge in Figure 5 are the quantity of activation and they are calculated with formula (1). Finally analyzing and interpreting the relationships between group performance and attributes of information flows.
The Operations of Graph
Create Minimum Spanning Tree
Finding the Shortest Paths
Finding the Second Shortest Path
Finding the K-Shortest Path
Finding the shortest path that must pass the two specified nodes
Traveling Salesman Problem
Dijkstra Algorithm
A* Algorithm
Greedy Algorithm
Branch and Bound
Dynamic Programming
Retrospective Algorithm
Enumeration
Floyd Algorithm
Kruskal Algorithm

Storage Structures of Graph
Adjacency List
Adjacency Matrix
Completed Graph

Calculating the second shortest path from V1 to V4

Figure 5. Portion of IIS-map with the quantity of activation

Drawing the initial map
Coding and segmenting information flows

The activated process
IIS-map with the quantity of activation

Figure 6. Screenshots of the analytic tool
Analytic Tool

We have developed the analytic tool that can draw the initial map, segment and transform information flows into sequences and compute attributes of information flows. Figure 6 shows the screen shots of the analytic tool.

Inter-rater reliability

In this study, two raters coded and segmented information flows of thirty groups and assessed all test papers independently. One was the first author and the other was the research assistant, a graduate student majoring in computer science. The coders were trained on how to segment information flows and assess test items. After the training sessions, each rater coded information flows of thirty groups and assessed ninety-two test papers independently. In order to confirm the reliability of the segmentation and assessing the test items, the percent agreement statistic proposed by Holsti’s (1969) was used to evaluate for inter-rater reliability. The percent agreement index was most common used method in content analysis and it reflected the number of agreements per total number of coding decisions (Rourke & Anderson, 2001). The total checklist included all items of the pretest and posttest and segmented information items of thirty groups. The two coders discussed and resolved all discrepancies. Reliability coefficient was calculated and results showed all values were above 0.9, regarded as an indication of excellent agreement.

Result

In order to test the hypotheses, the correlation analysis and linear regression analysis were conducted for group performance and attributes of information flows. Table 4 shows the mean and standard deviation of group performance, quantity of activation, average degree, average path length, density and degree entropy of the targeting knowledge network. Furthermore, the relationships between group performance and attributes of information flows were examined.

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group performance</td>
<td>29.734</td>
<td>14.931</td>
</tr>
<tr>
<td>Quantity of activation</td>
<td>254.716</td>
<td>128.914</td>
</tr>
<tr>
<td>Average degree</td>
<td>2.119</td>
<td>.088</td>
</tr>
<tr>
<td>Average path length</td>
<td>4.775</td>
<td>.391</td>
</tr>
<tr>
<td>Density</td>
<td>.087</td>
<td>.063</td>
</tr>
<tr>
<td>Degree entropy</td>
<td>1.328</td>
<td>.077</td>
</tr>
</tbody>
</table>

H1 assumed that the quantity of activation of the targeting knowledge network map could predict group performance. The result indicated that quantity of activation of the targeting knowledge network was significantly positively correlated with group performance ($r = .559, p = .001$). To examine the predictive validity of the quantity of activation on group performance, a linear regression analysis was conducted. The normal Q-Q plot was used to test normality of data. This test confirmed that the group performance variable had normal data. Consistent with hypothesis 1, the quantity of activation of the targeting knowledge network could predict group performance ($t = 3.565, \beta = .559, p = .001$). The quantity of activation could explain 28.8% of the total variance (Adjusted $R^2 = .288, F = 12.712$). This indicates that the quantity of activation of the targeting knowledge network map is the significant predictor. The main reason is that the quantity of activation of the targeting knowledge network map can measure the semantic relationships and deep structure of knowledge network map.

H2 assumed that the average degree of the targeting knowledge network could predict group performance. However, the data did not support this hypothesis: average degree was not correlated with group performance ($r = .063, p = .739$). So the average degree cannot predict group performance. The reason for this result is that average degree cannot represent the complexity of domain knowledge.
H3 assumed that the average path length of the targeting knowledge network could predict group performance. However, the data did not support this hypothesis: the average path length was not correlated with group performance ($r = .324, p = .081$). So the average path length cannot predict group performance. The major reason is that the average degree is a simple indicator for the size of knowledge network map.

H4 assumed that the density of the targeting knowledge network could predict group performance. However, the data did not support this hypothesis: the density was not correlated with group performance ($r = .233, p = .215$). So the density cannot predict group performance. The reason for this finding is that density cannot represent the deeper understanding of subject matter and semantic level.

H5 assumed that the degree entropy of the targeting knowledge network could predict group performance. However, the data did not support this hypothesis: the degree entropy was not correlated with group performance ($r = -.252, p = .180$). So degree entropy cannot predict group performance. The reason for this is that degree entropy only indicates the static typology structure of knowledge network map.

**Discussion**

The results of this study show that quantity of activation of the targeting knowledge network map can effectively predict group performance. This study also verifies that the IIS-map-based analysis method can analyze interactions in face-to-face collaborative learning. However, the results indicate that other four kinds of attributes including average degree, average path length, density and degree entropy of the targeting knowledge network cannot predict group performance. In fact these attributes can only reflect topological structures of the targeting knowledge network. They are static characteristics of the targeting knowledge network and cannot represent adequately complex knowledge structures. By contrast, quantity of activation can reflect the dynamic features of interactions. In formula (1) “d” denotes the number of the activated edge and it dynamically changes over time. So quantity of activation of the targeting knowledge network map can effectively predict group performance.

The implication of this result for collaborative learning is that teachers can predict which group is more successful by computing the quantity of activation of different groups. The higher the quantity of activation, the better group performance will be. The quantity of activation is one of the attributes of information flows, which can assess group performance. This attribute can be obtained without testing and it can be calculated by the analytic tool.

The IIS-map-based analysis method with the objective of seeking objectivity and reliability has a systematic approach to analyzing interactions. The collaborative learning system is considered as the complex information system. The information flow of this system is the main concern. IIS is an abstract generalization of an output sharing information set. Learners need to access information to acquire content knowledge and formulate hypotheses (Jonassen, 1999). So information is very important for knowledge gains. The IIS-map-based method aims to analyze the level of social construction of knowledge reflected by information flows. The approach of the IIS-map-based analysis method is different from previous studies. It sets aside some factors, such as attitudes, motivation, entry skills of learners, types of tasks, a variety of interactive strategies and various media or tools, which are classified as environmental factors of collaborative learning systems. On the contrary, information flows are identified as the analysis focus in this approach.

Stahl (2011a) believe that sequentiality and timing play an important role in how the postings are understood. The IIS-map-based analysis method highlights the temporal sequentiality by segmenting information flows in chronological order. It can fully convey the dynamic nature of interactions. So this method is of great significance for understanding the complicated interactions and knowledge construction.

The basic step of the IIS-map-based analysis method is to construct the IIS-map. Drawing the initial IIS-map needs to follow the accepted norm. Identifying the knowledge and constructing the correct relationships are the vital procedure. Of course, part of the initial IIS-map is likely revised when information flows of interaction process are inconsistent with the initial IIS-map. In addition, the test items that are used to test learning outcomes should be subjective items so as to reflect the cognitive process. The targeting knowledge need to be identified according to learning objectives that have been set up when designing collaborative tasks.
The sample of the IIS-map-based analysis method is not participants, but the targeting knowledge network map. One group’s interactive information flows generate one targeting knowledge network map. This study aims to analyze relationships between characteristics of targeting knowledge network map and learning outcome of the targeting knowledge network. This means that the IIS-map-based analysis method focuses on analyzing different attributes of the same object. We believe that this is more scientific than analyzing some attributes of different objects such as the learning outcomes of students and use of media.

Barron (2003) analyzed interactions of the twelve sixth grade triads by coding discourse into three kinds of speech acts (acceptance, discussion and rejection or ignorance). He drew a conclusion that less successful groups ignored or rejected correct proposals, whereas more successful groups discussed or accepted them. We examined the method of coding transcripts into speech acts using the data we collected. We selected fifteen more successful groups and fifteen less successful groups. The discourse transcripts were also coded into the same kinds of speech acts (acceptance, discussion and rejection or ignorance) by two coders. The reliability for acceptance, discussion and rejection or ignorance responses was 0.7, 0.75 and 0.68 respectively by computing Cohen’s (1960) Kappa. The result indicated that the more successful groups and less successful groups had no significant difference in accepting ($t = .258, p = 0.798$), discussing ($t = -1.371, p = 0.181$) and rejecting or ignoring ($t = .744, p = 0.463$). This result also proved that the method of coding transcripts into speech acts cannot distinguish between less successful groups and more successful groups. So Barron’s conclusion was not supported by our examination. When coding discourse we experienced the strong subjectivity and arbitrariness because coding was often based on subjective and borderline judgments and there was no objective reference. So it was easy to lead to the arbitrary interpretation and the low reliability conclusion.

In contrast to the approach of coding discourse into speech acts, the IIS-map-based analysis method is more objective and has a stronger predictive power. The essence of the IIS-map-based analysis method is to map information flows that are represented by natural language onto IIS-maps. The IIS-map serves as the frame of reference when information flows are segmented. This will help to diminish the subjectivity of coding and segmentation to a large extent so as to draw a reliable conclusion. The IIS-map with marks can distinctly reveal the characteristics of the collaborative learning system. The IIS-map-based analysis method can highlight the contribution of communicating analysis to learning analytics.

This study has several limitations. First, we only verify that the quantity of activation of the targeting knowledge network map can predict group performance. Other attributes of information flows will be explored in the future. Second, the sample is small and it originates from thirty group’s interactions. There is mainly one category of knowledge in this study. Following studies will explore more categories of knowledge and enlarge samples. Further, the follow-up study will explore how to represent and calculate emotions, attitudes and values on the IIS-map.

**Conclusion**

To sum up, the IIS-map-based method can support the analysis of interactions, and the quantity of activation of the targeting knowledge network map can predict group performance. Results of this study further indicate that the IIS-map-based method is an effective method for interaction analysis. The IIS-map-based method is a methodology for explaining a process-oriented account of group interactions. This method can provide insight into how collaborative learning takes place empirically. It is relatively objective because the IIS-map is the reference object of segmentation. Thus, the arbitrariness of coding discourse can be diminished in order to ensure the objectivity of segmentation of information flows. In addition, the quantity of activation can be calculated objectively.

There are multiple benefits to the IIS-map-based method as an analysis instrument. First, the IIS-map-based analysis method is useful for visualizing the interaction process and aiding the understanding and analysis of interactions. It is a useable and replicable instrument in collaborative learning. Second, an IIS map can unify data and has been used to support multiple analytic practices. The data represented by IIS-map are brief in revealing the characteristics of the collaborative learning system. Third, the IIS-map-based analysis method can directly examine the interaction process to predict group performance without traditional tests.
The IIS-map-based analysis method removes the subjectivity and arbitrary nature of value judgments when coding discussions and gives fully detailed analysis of interactions. This method will become a new analysis approach with high objectivity and strong feasibility. It will be widely applicable in the future.

Acknowledgement

This study was supported by the Twelfth Five-Year Plan of the National Educational Science in China (ECA110330). Many thanks for anonymous reviewers for their helpful comments and suggestions.

References


Appendix
Pre-test and post-test items

1. Please construct the minimum spanning tree of the graph (see Figure 1-1) by Prim's algorithm and Kruskal's algorithm and draw up the constructing process. What is the time complexity of the two algorithms respectively?

2. Figure 1-1. A undirected graph

2. The given vertex set and edge set of a weighted graph are respectively: \( V = \{1, 2, 3, 4, 5, 6\} \), \( E = \{<1, 2, 20>, <1, 3, 15>, <2, 1, 2>, <2, 5, 10>, <2, 6, 30>, <3, 2, 4>, <3, 6, 10>, <5, 4, 15>, <6, 5, 10>\} \).

   Please find the shortest path from vertex 1 to other vertices using Dijkstra's algorithm.

3. The given adjacency list of a graph is shown in Figure 1-2. Please solving the following problem:

   3.1 Find the adjacency matrix that corresponds to the adjacency list.
   3.2 Find the order of the depth-first traversal from the beginning of \( v_1 \).
   3.3 Find the spanning tree of the depth-first from the beginning of \( v_1 \).
   3.4 Find the order of the breadth-first traversal from the beginning of \( v_1 \).
   3.5 Find the spanning tree of the breadth-first from the beginning of \( v_1 \).

3. Figure 1-2. the adjacency list of a graph

4. Please design algorithms of touring a park where there are at least six scenic spots in order to provide services for visitors.

   4.1 How many kinds of storage structures can represent this problem? Explain the relative merits respectively.
   4.2 Provide the information consultant, including the name, code number and introduction of each scenic spot.
   4.3 If sewer pipes are reconstructed between each scenic spot, please design to how to reconstruct them at the minimum cost.
   4.4 If a visitor wants to tour from one of the scenic spots to another scenic spot, please design an algorithm to calculate the shortest path between them; In addition, please design the algorithm that calculates the second shortest path between two scenic spots.
   4.5 If a visitor wants to enter from the entrance to the park and leave from the exit and he/she wants to tour two famous scenery spots, please design the shortest path of this route.
   4.6 If a visitor sets out from the entrance to the park and tours each scenic spot only once, and then he/she leaves via the entrance, please design the shortest path for this route.
Dataset-Driven Research to Support Learning and Knowledge Analytics

Katrien Verbert¹, Nikos Manouselis², Hendrik Drachsler³ and Erik Duval¹

¹Dept. Computerwetenschappen, KU Leuven, Belgium // ²Agro-Know Technologies, Athens, Greece & University of Alcala, Spain // ³CELSTEC, Open University of the Netherlands, The Netherlands // katrien.verbert@cs.kuleuven.be // nikosmain@iee.org // hendrik.drachsler@ou.nl // erik.duval@cs.kuleuven.be

ABSTRACT

In various research areas, the availability of open datasets is considered as key for research and application purposes. These datasets are used as benchmarks to develop new algorithms and to compare them to other algorithms in given settings. Finding such available datasets for experimentation can be a challenging task in technology enhanced learning, as there are various sources of data that have not been identified and documented exhaustively. In this paper, we provide such an analysis of datasets that can be used for research on learning and knowledge analytics. First, we present a framework for the analysis of educational datasets. Then, we analyze existing datasets along the dimensions of this framework and outline future challenges for the collection and sharing of educational datasets.

Keywords

Learning and knowledge analytics, Datasets, Open science

Introduction

The need for better measurement, collection, analysis and reporting of data about learners has been identified by several researchers in the Technology Enhanced Learning (TEL) field (Siemens 2010; Romero et al. 2007; Duval 2011). This need has been translated into an emerging strand of research on learning and knowledge analytics (LAK), as reflected by a number of conferences and special issues in recent years (Siemens & Gasevic, 2011). Among others, the analysis of learner data and identification of patterns within these data are researched to predict learning outcomes, to suggest relevant resources and to detect error patterns or affects of learners. These objectives are researched to act upon needs of a variety of stakeholders, including learners, teachers and organizations. This is what drives major initiatives such as the US-based Learning Registry (http://www.learningregistry.org) to collect data and make them publicly available for research and application purposes.

Siemens (2010) defines learning analytics as “the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.” Contributions to the first conference on learning analytics and knowledge in 2011 indicate that information visualization, social network analysis and educational data mining techniques offer interesting perspectives for this emerging field. Whereas the specific techniques differ depending on context and the intended goals, the main objective of the approaches is to identify needs of target users and to support these needs using intelligent and adaptive systems.

Despite the recognition of the importance of LAK, the literature related to this topic is rather limited. Research on web analytics, search engines and recommender systems are excellent examples of how data gathering during an analytics cycle can be used to refine offerings to users (Elias, 2011). Whereas several recommender systems for learning (Manouselis et al., 2011), intelligent tutoring systems (Romero et al. 2007) and visual analytics systems (Govaerts et al. 2010) have been implemented for use in learning scenarios in recent years, many of these intelligent tools often stay in researcher hands and rarely go beyond the prototype stage (Reffay & Betheder, 2009). Among others, researchers have argued that the time needed by social science to validate prototypes is too long compared to the rate of technology innovation.

An important component to facilitate research in this area is the existence of extensive overviews of the available datasets that will provide researchers with a wide array of potential data sources to experiment with, as well as with an analysis of their properties that will help researchers decide about their appropriateness for their experiments. Such an overview is missing from LAK today, since only initial attempts have been made to document and study existing datasets (Drachsler et al., 2010). In this article, we extend our initial analysis (Verbert et al., 2011) of the datasets collected by the dataTEL Theme Team of the European Network of Excellence STELLAR (http://www.teleurope.eu/pg/groups/9405/datasetel/), in order to provide a more comprehensive overview of datasets for LAK research. The article makes four primary research contributions:
First, we present related initiatives that are collecting datasets and the needs and opportunities to make educational datasets available for LAK research. Second, we present a framework for the analysis of educational datasets. In particular, we present properties of educational datasets and LAK objectives that can benefit from the availability of such data. Third, we analyze existing datasets along the dimensions of our educational dataset framework. We also discuss existing research that has used these datasets for LAK related research. Finally, we present future challenges to enable the sharing and reuse of datasets among researchers in this field.

**Background**

In an increasing number of scientific disciplines, large data collections are emerging as important community resources (Chervenak et al., 2000). These datasets are used as benchmarks to develop new algorithms and compare them to other algorithms in given settings (Manouselis et al., 2010b). In datasets that are used for recommendations algorithms, such data can for instance be explicit (ratings) or implicit (downloads and tags) relevance indicators. These indicators are then for instance used to find users with similar interests as a basis to suggest items to a user.

To collect TEL datasets, the first dataTEL Challenge was launched as part of a workshop on Recommender Systems for TEL (RecSysTEL, Manouselis et al., 2010a) that was jointly organized by the 4th ACM Conference on Recommender Systems and the 5th European Conference on Technology Enhanced Learning in September 2010. In this call, research groups were invited to submit existing datasets from TEL applications. A special dataTEL Cafe event took place during the RecSysTEL workshop in Barcelona to discuss the submitted datasets and to facilitate dataset sharing in the TEL community.

Related work is carried out at the Pittsburgh Science of Learning Center (PSLC). The PSLC DataShop (Stamper et al., 2010) is a data repository that provides access to a large number of educational datasets derived from intelligent tutoring systems. Currently, more than 270 datasets are stored that record 58 million learner actions. Several researchers of the educational data mining community have used these datasets to predict learner performance.

The Mulce project (Reffay & Betbeder, 2009) is also collecting and sharing contextualized interaction data of learners. A platform is available to share, browse and analyze shared datasets. At the time of writing, 34 datasets are available on the portal, including a dataset of the Virtual Math Teams (VMT) project. This project investigated the use of online collaborative environments to support K-12 mathematics learning. These datasets have been used extensively by the Computer Supported Collaborative Learning (CSCL) community (Stahl, 2009).

Other efforts have been driven by fields studying child language acquisition. The CHILDES system (MacWhinney, 1996, 2007) helped realize much advancement in this field through sharing language-learning data. TalkBank (http://talkbank.org) is a follow-up project that is researching guidelines for ethical sharing of data, metadata and infrastructure for identifying available data, and education of researchers to the existence of shared data, tools, standards and best practices.

LinkedEducation.org is another initiative that provides an open platform to promote the use of data for educational purposes. At the time of writing, five organizations have contributed datasets. Available datasets describe the structure of organizations and institutions, the structure of courses, learning resources and interrelationships between people. In addition, various schemas and vocabularies are provided to describe the internal structure of an academic institution, discourse relationships, activity streams in social networks and educational resources. Such schemas and vocabularies offer interesting perspectives for the sharing and reuse of educational interaction data that is relevant for LAK research.

Several other initiatives are available that focus on providing the means to share datasets among researchers in a more generic way. DataCite.org is an organization that enables users to register research datasets and to assign persistent identifiers to them, so that datasets can be handled as citable scientific objects. The Dataverse Network (King, 2007) is an open-source application for publishing, citing and discovering research data. The network was established at Harvard University and is aimed to increase scholarly recognition for data contributions. Fact sheets of datasets are gathered from organizations and researchers are encouraged to make the data publicly available, if
possible. The Australian National Data Service (Treloar & Wilkinson, 2008) is a similar initiative in Australia that works on services to help researchers persistently identify and describe data.

In this paper, we analyze educational datasets that have been collected by dataTEL and related initiatives. We focus specifically on datasets that contain interaction/usage data of learners and that can be used for analytics’ research. In the next section, we present a framework for educational datasets that can be used to describe and analyze educational datasets. In addition, we discuss how work of related initiatives fits within this framework. Then, we analyze available datasets along the dimensions of this framework.

A Framework for Educational Datasets

In this section, we present a framework for the analysis of educational datasets. The framework is intended to address questions researchers might have about the potential usefulness of a dataset for their research purposes.

As illustrated in Figure 1, the framework constitutes three parts. **Dataset properties** describe the overall dataset, such as the application and the educational setting from which the data was collected. **Data properties** define at a finer grained level where data elements available, including action types such as downloads or selects and information about the learner and other entities involved. The third part of the framework defines a list of **objectives** of LAK research. These objectives are mapped to dataset and data properties in the next section to determine the potential usefulness of a dataset for LAK research purposes.

![Figure 1. Educational dataset framework](image)

**Dataset properties**

We recently (Drachsler et al., 2010) presented a specification of datasets that was used for the first dataTEL challenge. Among others, the specification includes information about the application in which the dataset has been collected, the educational setting, contact person, availability (open access or legal protection rules that describe how and when the dataset can be used), dataset collection method, dataset statistics, and pre-processing steps that have been applied to the data.
Related initiatives have also defined formats to package and describe datasets. As illustrated in Figure 2, a Mulce dataset is comprised of the following components:

- The instantiation component includes all interaction data, as well as user information.
- The learning design component describes the educational scenario.
- The research protocol describes the methodology of research with the dataset.
- The license component specifies dataset provider and user rights.
- The analyses component contains research outputs.

![Figure 2. Mulce format (Reffay & Betbeder, 2009)](image)

The PSLC DataShop project defines a specification for describing datasets that are derived from intelligent tutoring systems. The specification includes the project name, principal investigator, curriculum, collection dates, domain, application, description, hypothesis (e.g., “people who are required to use the tutor show less error on quizzes”), school, statistics and knowledge models of interactions. We discuss such interaction models in the next section. In addition, research papers with the datasets are referenced.

As illustrated in Table 1, there are many similarities between the specifications. Explicit information is indicated by “+” signs. This information constitutes explicitly articulated elements of the specifications. Implicit information is indicated by “(+)” signs and represents information that is implied or expressed as part of other elements. For instance, in the dataTEL specification, information about the domain or users can be described as part of the description of the application or environment, but no specific fields are provided for these elements. To date, the Mulce format provides the most comprehensive format for describing datasets. In addition to interaction data, the datasets incorporate a detailed description of the educational scenario in a learning design component and results of various analyses. Therefore, this specification provides the most interesting perspectives for describing educational datasets in a generic way.

**Data properties**

In addition to a format for describing datasets, there is a need to identify at a more fine-grained level of granularity which data elements are stored. Such information is essential to identify for which research purposes a dataset is useful. As outlined by Romero et al. (2007), the TEL field differs from the e-commerce analytics field in several ways. In e-commerce, the used data are often simple web server access logs or ratings of users on items. In TEL, many researchers use more information about a learner interaction (Pahl & Donnellan, 2002). The user model and the objectives of the systems are also different in both application domains (Drachsler et al., 2009).
A survey of existing TEL interaction data models has been presented in (Butoianu et al., 2010). Such models capture actions of users on resources, such as open/close, select/unselect or write actions, on resources. In addition, the context in which an action occurred can be captured, such as the current application the author is working with or her current task. The Atom activity stream RDF mapping of the LinkedEducation.org initiative provides such a model.
for actions of users in social networks. Vocabularies for actions, actors and objects involved and related contextual information are defined.

In addition to interaction models, learner models have been elaborated that describe several characteristics of learners. Brusilovsky and Millan (2007) identified the following categories based on an extensive analysis of the literature: knowledge levels, goals and tasks, interests, background and learning and cognitive styles. In addition, several models, standards and specifications have been elaborated to describe learning resources. The IEEE LOM and Dublin Core metadata standards are prominently used to describe learning resources, including general characteristics, such as title, author and keywords, technical and educational characteristics and relations between learning resources.

We integrated the various data categories and elements in Figure 3. We use this model in the remainder of this article to identify data elements in existing datasets. The model has been developed by synthesizing existing works on interaction data and context variables in the TEL field that were outlined above. It could be further refined by studying relevant theoretical frameworks, like the Activity Theory (Kaptelinin et al., 1995), which could help reorganize the various categories and elements. Future research work in this area is discussed in the last section.

**Learning and knowledge analytics objectives**

In order to provide guidance on the relevancy of datasets for LAK research, we identify a set of objectives that are relevant for LAK applications. We also outline existing research work in related research communities that, when interconnected, can provide substantial synergies to advance the emerging LAK field.

- **Predicting learner performance and modeling learners.** The prediction of learner performance and modeling of learners have been researched extensively by the educational data mining, educational user modeling and educational adaptive hypermedia communities. The objective is to estimate the unknown value of a variable that describes the learner, such as performance, knowledge, scores or learner grades (Romero & Ventura, 2007). Such predictions are for instance used by intelligent tutoring systems to provide advice or hints when a learner is solving a problem. Dynamic learner models are also researched to support adaptation in educational hypermedia systems (Brusilovsky & Millan, 2007).

- **Suggesting relevant learning resources.** Recommender systems for learning have gained increased interest in recent years. A recent survey of TEL recommender systems has been elaborated by Manouselis et al. (2011). These systems typically analyze learner data to suggest relevant learning resources, peer learners or learning paths.

- **Increasing reflection and awareness.** Several researchers are focusing on the analysis and visualization of different learning indicators to foster awareness and reflection about learning processes. These indicators include resource accesses, time spending and knowledge level indicators (Mazza & Milani, 2005).

- **Enhancing social learning environments.** Analysis and visualization of social interactions is researched to make people aware of their social context and to enable them to explore this context (Heer & boyd, 2005). In TEL, this is particularly, but not only, relevant for Computer Supported Collaborative Learning (CSCL) (Stahl, 2009), where the interactions with peer learners are a core aspect of how learning is organized. In CSCL, much research has focused on the analysis of networks of learners, typically with a social network analysis approach (Reffay & Chanier, 2003).

- **Detecting undesirable learner behaviors.** The objective of detecting undesirable learner behavior is to discover those learners who have some type of problem or unusual behavior, such as erroneous actions, misuse, cheating, dropping out or academic failure (Romero & Ventura, 2007).

- **Detecting affects of learners.** Researchers in TEL often refer to the affective states defined by D’Mello et al. (2007). These states are classified as boredom, confusion, frustration, eureka, flow/engagement, versus neutral. Among others, the detection of affects is researched to adjust pedagogical strategies during learning of complex material.

The objectives are highly interrelated. For instance, whereas research on affects and awareness and reflection has traditionally focused on an individual perspective, these objectives are also researched increasingly to enhance social learning environments.
Datasets for Learning and Knowledge Analytics

In this section, we present an analysis of datasets that can be used for a wide variety of LAK research purposes. We analyze the datasets along the dimensions of the dataset framework that we presented in the previous section.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Environment/Application</th>
<th>Collection Period</th>
<th>Statistics</th>
<th>Access rights</th>
<th>Educational Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendeley</td>
<td>Web portal</td>
<td>1 year</td>
<td>200,000 users 1,857,912 items 4,848,723 actions</td>
<td>Open access</td>
<td>Science</td>
</tr>
<tr>
<td>APOSDLE</td>
<td>PLE</td>
<td>3 months</td>
<td>6 users 163 items 1500 actions</td>
<td>Open access</td>
<td>Workplace Learning</td>
</tr>
<tr>
<td>ReMashed</td>
<td>PLE/Mail-up environment</td>
<td>2 years</td>
<td>140 users 960,000 items 23,264 actions</td>
<td>Legal protection</td>
<td>Computer science</td>
</tr>
<tr>
<td>Organic. Edunet</td>
<td>Web portal</td>
<td>9 months</td>
<td>1,000 users 11,000 items 920 actions</td>
<td>Legal protection</td>
<td>Agriculture</td>
</tr>
<tr>
<td>MACE</td>
<td>Web portal</td>
<td>3 years</td>
<td>1,148 users 12,000 items 461,982 actions</td>
<td>Legal protection</td>
<td>Architecture</td>
</tr>
<tr>
<td>Travel well</td>
<td>Web portal</td>
<td>6 months</td>
<td>98 users 1,923 items 16,353 actions</td>
<td>Open access</td>
<td>Various</td>
</tr>
<tr>
<td>ROLE</td>
<td>PLE</td>
<td>6 months</td>
<td>392 users 11,239 items 28,554 actions</td>
<td>Legal protection</td>
<td>Computer science</td>
</tr>
<tr>
<td>SidWeb</td>
<td>LMS</td>
<td>4 years</td>
<td>4,013,268 users 35,641 items 4,009,292 actions</td>
<td>Legal protection</td>
<td>Various</td>
</tr>
<tr>
<td>UC3M</td>
<td>Virtual machine</td>
<td>3 months</td>
<td>284 users 8,669 items 49,000 actions</td>
<td>Legal protection</td>
<td>Computer science</td>
</tr>
<tr>
<td>CGIAR</td>
<td>LMS</td>
<td>6 years</td>
<td>841 users 14,693 items 226,339 actions</td>
<td>Legal protection</td>
<td>Agroforestry</td>
</tr>
</tbody>
</table>

Table 2. Overview dataset properties

Dataset properties

Table 2 provides an overview of characteristics of available educational datasets, including the application from which data were collected, collection period, statistics and educational context or domain. The full description of the datasets is available on the portals that provide access to these datasets, including the dataTEL
Several dataTEL datasets have been collected from learning management systems (LMS). The UC3M dataset also collects data from a virtual machine that was used in a C programming course. The particularity of this dataset is that it records actions from several tools learners are using. The approach enables to collect a more comprehensive overview of learner activities, such as a learner searching for additional resources on the web. Such an approach is also researched under the prism of personal learning environments (PLEs), where data is tracked from learning environments that assemble relevant tools for course activities. Many other dataTEL datasets were collected from web portals that provide access to large collections of learning resources. Several other datasets are collected from intelligent tutoring systems (ITS) – including a large number of datasets from the PSLC DataShop initiative. We include the “Algebra 2008-2009” and “Bridge to Algebra” datasets that were used for the KDD 2010 Cup on educational data mining (https://pslcdatashop.web.cmu.edu/KDDCup/) in this analysis. In addition, the recommended datasets of the DataShop are analyzed. At the time of writing, 64 datasets are publicly available. Finally, many of the Mulce datasets contain data that were captured from forums, chat and email conversations between learners in collaborative learning settings.

The collection period varies from 10 days to 6 years. Several of the Mulce datasets capture data of group work during a specific learning activity. Datasets derived from learning management systems and web portals often capture data during a longer period of time, ranging from a couple of months to several years. Although a few datasets that are available capture data of a large number of users, many other datasets are more limited in size. Some datasets collect data of 1000 to 7000 users. Several other datasets capture data of a few learners only. These datasets are in some cases only a sample that the organization made available or in other cases datasets of a small number of collaborating users, such as the VMT and mce-copeas datasets.

Several datasets are openly accessible. For other datasets, legal protection rules apply. We obtained these datasets by sending a statement of our intended research purposes to the organization and then signed an agreement on the use of these data. All datasets contain data that is anonymized, so that it can no longer be linked to an individual.

Data properties

Table 3 presents a more detailed overview of the data elements that are included in the datasets. The datasets contain a diverse set of actions of users. These actions include attempts of learners on quizzes, search actions, selection, annotation, rating, creation or editing of resources. PSLC datasets derived from intelligent tutoring systems all include attempt actions on activities provided by the tutor. In some datasets, help requests are stored. The input provided by learners is sometimes further specified into select, write or create actions. The Mulce datasets capture social interactions – in most cases these constitute send and receive actions.

Explicit information about learners (or teachers) is stored in only a few datasets. The data is in most cases anonymized and little additional information about learners or teachers is stored. Some dataTEL datasets contain information about the language, interests, knowledge level or country of the user. Some DataShop datasets describe the gender and knowledge level of the learner, including her past grades. The mce-copeas dataset divides learners in three groups according to their knowledge level (beginner, medium, expert). Information about country, age, language and gender is often provided in Mulce datasets.

Information about resources is available in more datasets. The information ranges from an identifier of the resource to detailed descriptions that include educational characteristics such as duration, minimum age, maximum age and resource type, technical characteristics and annotations such as tags and comments. Such metadata are often provided in dataTEL datasets that were captured from learning repository portals. In the DataShop datasets, educational information such as average duration and required skills are sometimes provided. In addition, compositional relations are provided that define a hierarchy of units and sections. Social relations between learners collaborating are stored in the Mulce and some dataTEL datasets.

Additional context information is also stored. Several datasets provide timestamp information. The duration of an action is stored explicitly as a time interval in the DataShop and some LMS datasets. Such information is valuable to
calculate the difference of estimated durations, described in resource metadata, and the time the learner needed in practice to complete an activity. Other contextual information is not often available. In datasets that contain data of multiple tools and services, information about the application from which an action was triggered is included.

Table 3. Overview data properties

<table>
<thead>
<tr>
<th>action type</th>
<th>dataTEL</th>
<th>PSLE dataShop</th>
<th>Mulce</th>
</tr>
</thead>
<tbody>
<tr>
<td>attempt</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>create/delete</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>write/edit</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>select/unselect</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>search</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>send/receive</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>register/unregister</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>open/close</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>add/remove</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>tag/annotate</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>rate</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>request help</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>save/download</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>login/logout</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

| learner/teacher      |         |               |       |
| id                   | +       | +             | +     |
| country/school       | +       | +             | +     |
| gender               | +       | +             | +     |
| age                  | +       | +             | +     |
| languages            | +       | +             | +     |
| knowledge level      | +       | +             | +     |
| interest             | +       | +             | +     |
| goals/tasks background|        |               |       |

| resource             |         |               |       |
| general              | +       | +             | +     |
| technical            | +       | +             | +     |
| educational          | +       | +             | +     |
| relation             | +       | +             | +     |
| classification/annotation | +     | +             | +     |

| context              |         |               |       |
| timestamp            | +       | +             | +     |
| duration             | +       | +             | +     |
| application          | +       | +             | +     |

| result               |         |               |       |
| error message        | +       | +             | +     |
| success/failure      | +       | +             | +     |
| score/grade          | +       | +             | +     |
| content (rating value, tewe. etc.) | +       | +             | +     |

Finally, the result of actions, such as correct or incorrect attempts, rating values or error messages, is stored. In addition, some datasets contain the grade a learner obtained for an activity or course. We elaborate in the next section how such data can be used for LAK research.
Usefulness of available datasets for LAK related research objectives

Prediction of learner performance and discovering learner models

Several datasets are available that can support research on prediction of learner performance and discovery of learner models. Among others, such predictions are researched to provide advice when a learner is solving a problem (Romero et al., 2007). Datasets from intelligent tutoring systems that capture attempts of learners provide a rich source of data to estimate the knowledge level of a learner. Some datasets derived from LMSs contain data on the number of attempts and total time spent on assignments, forums and quizzes. Romero et al. (2008) compared different data mining techniques to classify learners based on such LMS data and the final grade obtained for courses. Datasets that have been captured from PLEs offer interesting perspectives to elaborate such research in open learning environments. In addition, many datasets are suitable to identify interests of users based on resource accesses.

Several researchers have already experimented with the datasets outlined above to predict learner attributes. The “Algebra 2008-2009” and “Bridge to Algebra” datasets were used in the KDD Challenge 2010. Participants were asked to learn a model from past learner behavior and to predict their future performance. The winners of this competition combined several educational data mining techniques. Cen et al. (2007) performed a learning curve analysis with the “Geometry Area” dataset. They noticed that while learners were required to over-practice some easy target skills, they under-practiced harder skills. Based on this observation, they created a new version of the tutor by resetting parameters that determine how often skills are practiced. References to other studies with these datasets are available on the DataShop portal.

Suggesting learning resources

Several dataTEL datasets contain explicit relevance indicators in the form of ratings that are relevant for research on recommendation algorithms for learning. In addition, implicit relevance indicators, such as downloads, search terms and annotations, are available that can be used for such research. If time interval data is available, the data might be suitable to extract reading times in order to determine the relevancy of a resource. In addition, such datasets are useful to analyze information about sequences of resources as a basis to suggest learning paths.

Manouselis et al. (2010b) used the Travel well dataset to evaluate recommendation algorithms for learning. Similar experiments have been reported in (Verbert et al., 2011). In this study, the Mendeley and MACE datasets were also used. Although still preliminary, some conclusions were drawn about successful parameterization of collaborative filtering algorithms for learning. Outcomes suggest that the use of implicit relevance indicators, such as downloads, tags and read actions, are useful to suggest learning resources.

Increasing reflection and awareness

Several datasets are useful for analysis and visualization of different learning indicators to foster awareness and reflection about learning processes. In addition to indicators about the knowledge level of learners, several datasets contain indicators of the time learners spend on learning activities – such as the PSLC DataShop datasets.

Other datasets contain timestamp information that can be used to derive indicators of the time users were active. dataTEL datasets were for instance used to obtain such indicators as a basis to support awareness for teachers (Govaerts et al. 2010). A visualization of these indicators applied to the ROLE dataset is illustrated in Figure 4. Evaluation results indicate that the perceived usefulness for teachers is high. The MACE dataset has been used for research on reflection and awareness of resource accesses. The Zeitgeist application (Shmidz et al., 2009) gives users insight into which learning resources they accessed, how they found them and which topics have been of interest to them (see Figure 5).
Figure 4. Visualization of time indicators (Govaerts et al., 2010)

Figure 5. MACE Zeitgeist (Shmidz et al., 2009)

Enhancing social learning environments

Several Mulce datasets are useful for research on collaborative learning. The datasets have been captured from chat tools, forums or email clients. Such data can be analyzed to predict and advice on learning in group work. Datasets that have been captured from LMSs often capture messages within course forums. Some of the PSLC DataShop
Datasets capture collaborative activities with intelligent tutoring systems, including the “Electric Fields – Pitt” dataset.

Several datasets have already been used to support research on enhancing social learning environments. Research with the “Electric Fields – Pitt” dataset suggests that asking learners to solve problems collaboratively with an intelligent tutoring system is a productive way to enhance learning from an ITS (Hausmann et al., 2008).

Several Mulce datasets have been used for research on collaborative learning (Stahl, 2009). Among others, the datasets have been used to understand mathematical ideas and reasoning in chat by learners, interaction mechanisms used by online groups to sustain knowledge building over time and the measurement of cohesion in collaborative distance learning. Evaluation studies showed that such analysis, when embodied in visualization tools (see Figure 6), can efficiently assist the teacher in following the group collaboration (Reffay & Chanier, 2003). These analyses were used to highlight isolated people, active sub-groups and various roles of the members in group communication. The mce-copes dataset has been used to research the influence of synchronous communication during online collaborative writing activities (Ciekanski & Chanier, 2008). Several other studies are documented on the Mulce portal.

![Figure 6. Matrix and graphical representation of e-mail exchange (Reffay & Chanier, 2003)](image)

**Detecting undesirable learner behaviors**

Datasets derived from PLEs provide a rich source of data to detect unusual behavior, as these datasets record actions of learners with several tools they were using during the classes. Data from LMSs can be used to detect potential dropouts when learners are no longer active. ITS datasets are also suitable for research on unusual behavior. Baker et al. (2006) found that learners who were “gaming the system” (i.e., fast and repeated requests for help to avoid thinking) had the largest correlation with poor learning outcomes.

![Figure 6. Pattern visualization of UC3M dataset (Scheffel et al., 2011)](image)
Scheffel et al. (2011) used the UC3M dataset to identify key actions from observed learning behavior. The authors employed data mining techniques to extract frequent patterns of actions. These patterns were visualized to support teaching activities. For instance, the pattern illustrated in Figure 7 points to development flows in which for each compilation students opened a file and closed it again before compiling. According to the teaching staff, such actions translate into a significant increase in development time and should be corrected.

Detecting affects of learners

Some datasets are suitable for research on the detection of affects and motivational aspects. For instance, PSLC DataShop datasets can be used to extract motivational aspects by comparing the time a learner spends on a learning activity in an ITS with the expected or average time of other learners. The use of emoticons and affective words is researched in social interaction datasets (Reffay et al., 2011). Prominent research in building a user model of affect from real data has been conducted by Conati and Maclaren (2005).

Ongoing research with the UC3M dataset is focused on the detection of affects of learners, such as frustration and (dis-)engagement. Based on an analysis of sequences of actions, such as a sequence of error messages of a debugger and successful compilations, information is deduced about potential engagement or frustration.

Conclusion and future challenges

In this article, we have presented an overview of datasets that can be used for exploratory research on LAK. Several datasets have been identified and analyzed along the dimensions of our educational dataset framework. Our analysis indicates that an initial collection of interesting datasets is already available. These datasets have been used in a successful way by several researchers, for diverse research purposes that are relevant for LAK analytics. The analysis provides researchers with an overview of available datasets, as well as their properties to help decide about their appropriateness for their experiments. In addition, the analysis sheds some light on what data to track for LAK research. The datasets were collected by our dataTEL and related initiatives, with the common objective to make educational datasets available to researchers.

Nevertheless, our endeavors to collect and share datasets for research remain quite challenging. A first challenge is related to privacy rights and licensing of educational data. Although an enormous amount of data has been captured from learning environments, it is a difficult process to make these data available for research purposes. The issue of usage rights/licensing needs to be solved from two perspectives. From a user perspective, learners need to be informed and grant permission to collect their data and make it available for research purposes. Also the organization or provider of these data needs to agree with collecting and sharing these data. For instance, researchers have in some cases collected datasets by crawling data from websites and then found out that they were not allowed to do so.

The collection of the UC3M dataset is a good example of how data can be made available for research purposes. In a first stage, learners were informed about which data were collected during their course activities and signed an agreement that the data could be used for research purposes, as is required by the Spanish law on data protection and privacy. In a second stage, researchers signed an agreement on the use of the dataset for research purposes. In order to facilitate the collection of educational datasets, it is important to share such practices and to contribute to the documentation of legal data protection and privacy laws. In addition, providing guidelines for anonymization of data and creating incentives for researchers and organizations to share their datasets is important. The first dataTEL challenge requested submissions of dataset fact sheets in a pre-defined format so that required properties could be well documented. The approach is similar to work of the DataVerse Network (King, 2007).

A second challenge is related to the heterogeneity of educational datasets. The lack of a standard representation for interaction data within datasets prevents the sharing and reuse of data across systems. In addition, when a custom data format is not well documented, it may be difficult to assess the meaning and usefulness of data elements that are stored. To address this issue, the development of a standard representation for learner interaction data will be taken up by a working group of the CEN Workshop on Learning Technologies (WS/LT).
This challenge sets the context of a third challenge, the identification of relevant data about learners and other entities involved for LAK research. Whereas we were able to identify some data elements that can be used as input for the CEN WS/LT working group, additional research is required to identify a broader set of elements that are useful for LAK research. In addition, the model could be further refined by studying relevant theoretical frameworks, like the Activity Theory (Kaptelinin et al., 1995), which could help us reorganize the various categories and elements according to well-established frameworks.

A fourth challenge is the development of data sensors to collect data. Data that is tracked within learning management systems provides a good basis for exploratory research on learning and knowledge analytics. However, learners often use a wide variety of tools and services in addition to a traditional LMS. Ongoing work within the ROLE project (http://www.role-project.eu) is focused on collecting data from PLEs that aggregate several tools and services. The approach is inspired by wakoopa (http://social.wakoopa.com) and rescuetime (http://www.rescuetime.com) that install tracker tools on the machine of a user and automatically record all activities.

In addition, data sharing and reuse in the educational field needs further research to explore whether the context and scope of the dataset collection can significantly affect its potential reuse. This is a requirement to keep in mind, as relevant discussions on productive multivocality in CSCL have indicated that there is a possibility that data collected using one theoretical framing may be unsuitable for analysis under another framing. The issue also indicates that clear descriptions of datasets along various dimensions describing such properties, as provided by the Mulce and DataShop specifications, are important for exchanging and possibly reusing datasets that have been collected in various settings.

Finally, the interconnection of several efforts in the area of educational dataset collection is key to advance this work. dataTEL has recently been accepted as an EATEL Special Interest Group (SIG). The dataTEL SIG aims to bring together existing efforts in the area of dataset collection and sharing. The SIG will organize yearly workshops and three monthly virtual meetings. With the organization of these events, we hope to enable collection and sharing of datasets on a much larger scale than available today.

**Acknowledgements**

Part of this work was supported by the dataTEL Theme Team of the FP7 STELLAR Network of Excellence (no. 231913). We thank all members involved, and in particular Martin Wolpers, Miguel-Angel Sicilia, Stefanie Lindstaedt and Riina Vuorikari, for their contributions to the organization of dataTEL events. Katrien Verbert is a Postdoctoral Fellow of the Research Foundation-Flanders (FWO). The work of Nikos Manouselis has been funded with support by the EU project VOA3R-250525 of the CIP PSP Programme (http://voa3r.eu).

**References**


Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan

Leah P. Macfadyen* and Shane Dawson1,2

1Arts ISIT, Faculty of Arts, The University of British Columbia, Vancouver, BC, Canada // 2Faculty of Education, University of Wollongong, Australia // leah.macfadyen@ubc.ca // sdawson@exchange.ubc.ca

*Corresponding author

ABSTRACT

Learning analytics offers higher education valuable insights that can inform strategic decision-making regarding resource allocation for educational excellence. Research demonstrates that learning management systems (LMSs) can increase student sense of community, support learning communities and enhance student engagement and success, and LMSs have therefore become core enterprise component in many universities. We were invited to undertake a current state analysis of enterprise LMS use in a large research-intensive university, to provide data to inform and guide an LMS review and strategic planning process. Using a new e-learning analytics platform, combined with data visualization and participant observation, we prepared a detailed snapshot of current LMS use patterns and trends and their relationship to student learning outcomes. This paper presents selected data from this “current state analysis” and comments on what it reveals about the comparative effectiveness of this institution’s LMS integration in the service of learning and teaching. More critically, it discusses the reality that the institutional planning process was nonetheless dominated by technical concerns, and made little use of the intelligence revealed by the analytics process. To explain this phenomenon we consider theories of change management and resistance to innovation, and argue that to have meaningful impact, learning analytics proponents must also delve into the socio-technical sphere to ensure that learning analytics data are presented to those involved in strategic institutional planning in ways that have the power to motivate organizational adoption and cultural change.

Keywords

Learning management system (LMS), Virtual learning environment (VLE), Learning analytics, Strategic planning, Student engagement, Change management, Institutional culture

Introduction

The promise of learning analytics

Learning analytics employs sophisticated analytic tools and processes in investigation and visualization of large institutional data sets, in the service of improving learning and education (Brown, 2011; Buckingham Shum & Ferguson, 2011). Building on the demonstrated strategic advantages of “business analytics” in the corporate world, learning analytics also draws on the related fields of web analytics, academic analytics (Goldstein & Katz, 2005), educational data mining (see Romero & Ventura (2010) for review) and action analytics (Norris, Baer, Leonard, Pugliese, & Lefrere, 2008) to support decision-making and strategic planning in academic settings. “Academic analytics” approaches are typically applied in educational settings to address administrative and operational concerns such as “advancement/fundraising, business and finance, budget and planning, institutional research, human resources, research administration, and academic affairs” (Fritz, 2011). Projects undertaken under the auspices of “learning analytics” extend the potential of analytics to the level of individual learning, by selecting, capturing and interpreting data on teaching and learning activities, with the goal of improving teaching and learning outcomes. Institutions and senior administrators are key users and stakeholders, and enhancement of institutional decision-making processes and resource allocation are core objectives (Romero & Ventura, 2010). In the postmodern context of constant and dynamic change in higher education and technological innovation, then, learning analytics offers higher education institutions a valuable tool in their ongoing efforts to select actions that are “achievable within the capacity of the organization to absorb change and resource constraints” (Kavanagh & Ashkanasy, 2006).

The importance of strategic investment in learning technologies and e-learning

In this paper, we present a learning analytics case study of LMS implementation and use in a large research-intensive university that routinely ranks within the top five in national magazine league tables (Macleans, 2010). The institution achieves high annual scores in the presage variables compiled in such rankings: institutional resources,
research funding and reputation. We know from meta-analytic studies of decades of available data, however, that the quality of education offered by an institution is not predicted by the size of institutional budgets, numbers of or dollar values of research awards, or even by measures such as student: faculty ratios or “hours spent in class.” Instead, the best institutional predictors of educational gain are “measures of educational process: what institutions do with their resources to make the most of whatever students they have” (Gibbs, 2010, p. 2). Citing a major 2004 study (Gansemer-Topf, Saunders, Schuh, & Shelley, 2004), Gibbs (2010) argues that the feature that distinguishes effective institutions from less effective schools is their strategic use of available funding to support “a campus ethos devoted to student success” (p. 14). In other words, decision-making processes relating to organization of institutional resources – human and material – and planning for more effective use of existing resources are a critical feature of excellent institutions.

Within the teaching context, Gibbs argues that the most significant predictors of educational gain “concern a small range of fairly well understood pedagogical practices that engender student engagement” (2010, p. 5). At least a decade of research and writing has demonstrated that learning technologies, when used appropriately, can help educators adopt the “seven principles of good practice in undergraduate education” (Chickering & Gamson, 1987) and improve the overall quality of an institution’s educational delivery (Chickering & Ehrmann, 2002). Moreover, recent work in the field of learning analytics has demonstrated that the communicative affordances of ICTs and learning management systems (LMSs) can increase student sense of community (Dawson, 2006) support learning communities and enhance student engagement (Dawson, Burnett, & O'Donohue, 2006; Dawson, Heathcote, & Poole, 2010; Macfadyen & Dawson, 2010).

The teaching climate within higher education is becoming increasingly complex. Student enrollment numbers continue to rise (Patrick & Gaële, 2007) and universities are catering to an increasingly diverse student body (living far from campus, studying part-time, returning to education after a long break or juggling the demands of study with career or family life (OECD, 2008; Twigg, 1994)). In this context, learning tools that support and enhance student engagement with peers, instructors and learning materials have become essential enterprise resources. It should be no surprise then, that, like 93% of US-based higher education institutions (Campus Computing, 2010), the institution in this case study has invested heavily in the campus-wide implementation of a web-based LMS since the late 1990s. The institution also hosts and supports a number of additional learning technology platforms (e.g., WordPress, MediaWiki), and subscribes to others (e.g., Turnitin). The LMS is therefore embedded within a wider network of platforms and systems involved in supporting the teaching and learning mission, and is viewed as a core component of the university’s teaching infrastructure.

The catalyst for change

Given this university’s substantial investment in an institutional LMS, and the ever-evolving market in LMSs and learning technologies, as well as shifting economic conditions (Campus Computing, 2010), strategic decision-making and forward planning regarding technology choices and related resource allocation are clearly of the essence. Reviews of available learning technologies have routinely been undertaken as part of the institution’s standard quality assurance practices. These reviews have aimed to evaluate the current state of use, to ensure that the suite of adopted technologies reflect the broader learning and teaching mandate and to ensure that the university is deploying its limited resources most effectively to support learning and teaching. In addition, in 2010, a further catalyst for a new LMS review was the LMS vendor’s announcement that the current LMS product would not be supported after 2013. Together, these conditions generated the necessary impetus for the next round of institutional review, with the goal of selecting as the next enterprise LMS the product that would best support the university’s teaching and learning goals.

In his well-established eight-step change model, Kotter (1996) notes that the critical first step in effectively managing change is one that creates a sense of urgency and the necessary levels of motivation required for sustaining the change process. This step involves a careful examination of the current context, to allow identification of potential “threats” and opportunities, and envisioning of future scenarios. In this light, the application of learning analytics focused on the current state of LMS integration presented an opportunity to understand the specific teaching and learning context, and develop a strategic vision and operational pathway for continual improvement.
Employing e-learning analytics to undertake a current state analysis

Gathering as much data as possible regarding current LMS usage across a university is no small feat, and is particularly challenging in such a highly decentralized institution as the one under study here, which comprises numerous Faculties/Divisions. Some LMSs do capture and store large amounts of course and user activity and interaction data. Until recently, however, investigators have only been able to access, aggregate, analyze, visualize and interpret this data via slow and cumbersome manual processes. While the majority of commercial and open source LMS are rapidly recognizing the importance for integrating sophisticated learning analytics, the associated reporting functionality is still largely under development (Dawson, McWilliam, & Tan, 2008; Mazza & Dimitrova, 2007).

To overcome the challenge of poor analytics functionality in the current LMS, the university considered here has partnered with an analytics software company to customize and implement an analytics reporting tool that allows extraction, analysis and dis/aggregation of detailed information about uptake and use of the enterprise LMS. We made use of this analytics platform to carry out the requested “current state analysis” of LMS usage, and to seek answers to questions about the extent and complexity of LMS adoption. This analysis was undertaken with the goal of informing and guiding the institution’s campus-wide strategic planning process for learning technology and LMS integration.

The availability of the new e-learning analytics platform allowed us to undertake, on the university’s behalf, the most comprehensive examination of its LMS use to date. The process revealed, and will continue to reveal, an array of LMS use patterns and practices, finally allowing the university to “know itself” in terms of learning technology uptake and integration in the service of teaching and learning. We hoped that this e-learning analytics exercise would provide compelling data that would generate the sense of urgency necessary for motivating broad scale institutional change associated with learning, teaching and technology. Through participant observation in the review and planning process we were able to investigate the degree to which the e-learning intelligence revealed influenced institutional decision-making.

In this paper we present selected examples of data from the analysis. More critically, we discuss the reality that the data developed in this e-learning analytics process did not significantly inform subsequent strategic decision-making and visioning processes, and consider some of the factors that may have limited its impact.

Approach and tools

Ethics and privacy

There are very real concerns about ethics and information privacy issues relating to the collection, analysis and dissemination of data on student, faculty and staff online activity and on student achievement and demographics. For this reason, our approach is informed by the institution’s policies on research involving human subjects, and the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (“TCPS”) (Government of Canada, 2010). These policies mandate an ethical review process for research projects that involve human subjects. Furthermore, data that is gathered through institutional research is subject to the provisions of the Freedom of Information and Protection of Privacy Act (Government of British Columbia, 2012) (whereas data generated in the course of “traditional academic research” is under the control of the individual faculty members involved and exempted under section 3(1)(e) of the Act.). This research complies with all stated policies and in accordance with FIPPA, our data is protected. Data access is limited to a small number of research investigators; raw data is maintained on secure data servers only; and all individual identifiers are removed from any disseminated data or analysis.

Selection of data

LMS usage data was mined for a single academic year (2009-2010). Only data for credit-bearing courses was examined (LMS adoption by the institution’s various continuing education units for non-credit programming, and usage by non-teaching units, was excluded, as well as numerous non-credit LMS-based online “training and
orientation” modules created by Faculties and support units). Data on the total number of courses offered across the institution in the 2009-2010 academic year was prepared by manual analysis of course section listings exported from an institutional database listing all sections available for student registration in 2009-2010. Sections listed as summer courses, registration placeholders for overseas exchange courses, or registration placeholder course numbers for students completing Masters or Doctoral theses were removed, before determining section counts and statistics. The following section outlines some of the analyses that were undertaken.

Analytics and data visualization tools

E-learning data presented in this paper was extracted, collated and analyzed using an e-learning analytics platform based on MicroStrategy Business Intelligence software that allows development of customized reports investigating user, course, Department, Faculty and institutional metrics during time periods of interest. User interaction data within the institutional network can be readily captured, categorized, and analyzed. These large data sets can be further interrogated to identify patterns of user-behaviour that can inform teaching and learning practice. The software makes use of a pre-built enterprise data warehouse, optimized for use by educational organizations, with common, conformed dimensions, enabling cross-platform intelligence.

At the institution under study here, the analytics platform has been configured to pull course identifier data and student grades from the institution’s student information system, LMS data from LMS tracking and meta-data (captured on production servers). Data is presented to end users via a web interface with extensive report authoring, ad hoc querying, data analysis and report distribution capabilities. Selected data was also visualized using Tableau Desktop 6.1 data visualization software. (For more information http://www.tableausoftware.com/)

Participant observation

To investigate the subsequent impact of our e-learning analytics reporting on institutional decision-making processes in relation to LMS selection and learning technology planning, we undertook a longitudinal participant observation process (Douglas, 1976). In participant observation, investigators are also subjects. This qualitative research methodology typically involves direct observation, participation in the life of the group, collective discussions, and analysis of documents produced within the group. It is usually undertaken over an extended period of time, ranging from several months to many years.

The institution in question hosts a standing advisory committee on learning technologies that comprises at least 35 representatives from across its academic, information technology and learning technology units, and is jointly chaired by senior figures in academic affairs and information technology. We participated in and observed the activities and collective discussions of the committee over a period of approximately 18 months, during which time it was tasked with evaluating current usage of the institutions LMS and other tools, and development of a vision, roadmap and plan for the institution’s next generation learning technology environment. We also undertook review and analysis of public and private process documents developed by the committee. The lead author of this work participated with ‘observer status’ only, and played no role in decision-making processes.

Selected findings

Institutional data indicated that in 2009-2010 academic year, a total of 18,909 course sections were offered (of which 14,201 were undergraduate course sections). This total includes 388 distance learning sections, of which 304 (1.6% of total sections) were offered in fully online format, and 84 in print-based format.

Numbers of courses and students using the enterprise LMS

Assessment of the proportion and characteristics of course sections implementing the LMS provides a sound indication of broad-scale institutional adoption rates and overall diversity of adoption across year levels and class sizes. In this instance, 21% (3,905) of all course sections had an associated LMS course site.
Based on institutional staffing and student enrollment figures for 2009-2010:

- 80.3% of all students were enrolled in at least one LMS-supported course during the 2009-2010 academic year (total student enrollment: 52,917)
- Most LMS-supported sections (61%) were employed for medium-sized course sections of 15-79 students. A further 22% of sections were employed for large classes of 80+ students.
- 1,118 instructors or roughly 30% of all teaching staff used the LMS for instructional purposes (total teaching staff of 3,061, including part-time and full-time Professors; Associate, and Assistant Professors; lecturers; instructors; and clinical, visiting, adjunct and emeritus Faculty).

The institution’s LMS is currently used by courses across all year levels (1st-4th year undergraduate courses, as well as graduate level courses), with roughly 14% of lower level course sections and graduate course sections, and 25% of upper level course sections making use of the LMS. Across the undergraduate years, numbers of unique student users are similar (ranging from 12,000-19,000 unique student users), demonstrating that in upper level (3rd and 4th year courses) the LMS is, on average, being used to support smaller course sections than at the lower level. While fully online courses represent only 1.6% of course section offerings in 2009-2010, 4,661 students or 11% of the total enrollment completed at least one fully online course during this period.

### LMS user time online

User time online within LMS-supported course sites varied immensely by user role (designer, instructor, teaching assistant, student), by Faculty, by Department, and by course mode (fully online versus LMS-supported). Table 1 shows comparative average user time online per term for LMS-supported and fully online courses.

<table>
<thead>
<tr>
<th>Role</th>
<th>LMS-supported courses</th>
<th>Online courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designer</td>
<td>6 ± 15</td>
<td>23 ± 16</td>
</tr>
<tr>
<td>Instructor</td>
<td>2 ± 6</td>
<td>17 ± 26</td>
</tr>
<tr>
<td>Teaching Assistant</td>
<td>7 ± 27</td>
<td>11 ± 23</td>
</tr>
<tr>
<td>Student</td>
<td>9 ± 9</td>
<td>41 ± 26</td>
</tr>
</tbody>
</table>

Average user time figures mask real variation between Faculties, Departments and even individual course sections. Students in LMS-supported courses in the Faculty of Arts, for example, spent an average of 7 ± 6 hours per course section using LMS-based course resources, while students in the Faculty of Agriculture spent an average of 16 ± 15 hours online per course section.

Similarly, instructor time online varied tremendously, even when courses were taught in a fully online modality. Examination of instructor data for fully online courses shows a range of instructor time online from 61 ± 261 hours per section at the high end, to 5 ± 4 hours per section at the low end in the 2009-2010 academic year.

### What learners are doing online

Measures of “average time online” using an LMS is a crude indicator of student (or instructor) time investment in teaching and learning. In order to further unpack what students are doing while logged in to LMS-based course sites, we investigated data on LMS tool use.

The current institutional LMS offers instructors and students a range of tools for presenting learning materials, communication, collaborative work, assessment and administrative tasks. In addition, a number of web-based products and services (Turnitin, MediaWiki, the Wimba suite of tools) offer “plugins” – known as Powerlinks – that allow their integration into an LMS-based course. Assessment of LMS tool “presence” in LMS-supported course sections during the period of interest shows that the standard suite of LMS tools are typically implemented (i.e., available for use) as well as a number of Powerlinks (data not shown). However, a more nuanced representation of LMS tool use is provided in Figure 1, which illustrates average actual tool use time per student (measured in minutes) for all available tools and Powerlinks in the 2009-2010 session.
Figure 1. Student usage of LMS tools shown as minutes of use time per student enrolled in LMS-supported course(s), 2009-2010

Figure 2. Overall composition of the institution’s LMS-based course content, represented as relative numbers of each file type
To explore the nature of actual learning materials (i.e., course content) we investigated which file types are contained in the entire LMS course content database. The diversity and proportion of file types is represented in Figure 2. This data can provide insight into the types of learning and teaching approaches adopted for a particular course, or can allow a more generalized assessment across a Department or Faculty.

![Figure 2. Diversity and proportion of file types in the LMS course content database.](image)

The aggregation of tool use data based on tool “purpose” provides an effective method for interpreting the broad pedagogical intention of online learning materials and activities. Dawson (Dawson et al., 2008) has previously proposed that LMS tools can be broadly organized into four categories representing the core activities within LMS-supported and online courses:

- Engagement with learning community
- Working with content
- Assessment
- Administrative tasks

**Table 2. LMS tools assigned to “learning activity categories”**

<table>
<thead>
<tr>
<th>Administration</th>
<th>Assessment</th>
<th>Content</th>
<th>Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>application login</td>
<td>assessment</td>
<td>search</td>
<td>Voice Email</td>
</tr>
<tr>
<td>compiler</td>
<td>assignments</td>
<td>media-library</td>
<td>Live Classroom</td>
</tr>
</tbody>
</table>
Figure 4. Correlation between student achievement and selected LMS tool use frequency in LMS-supported course sections
This categorization offers a useful approach for interpreting LMS tool use data, especially in light of increasing evidence that student engagement with peers in a learning community has the strongest positive effect on learning success (Astin, 1993; Light, 2001; Macfadyen & Dawson, 2010; Tinto, 1998). Table 2 outlines our assignment of currently available LMS tools into these learning activity categories. Figure 3 represents average learner time using each tool category in LMS-enabled course sections proportionately for each Faculty, regardless of absolute time use figures. This allows easy comparison of relative tool category use time per student between Faculties.

LMS tracking data and student grade data for 95,132 undergraduate student enrollments in LMS-supported courses was merged, visualized and analyzed using Tableau 6.1. Student grades were binned into deciles and best fit lines determined. Correlation coefficients for the selected LMS activities shown here with binned student final grade are as follows: number of discussion messages posted, $r = .83$, $p<.01$; number of discussion messages read, $r = .95$, $p<.0001$, number of discussion replies posted, $r = .94$, $p<.0001$; number of content pages viewed (0.89, $p<.001$); number of visits to the “My Grades” tool (0.93, $p<.0001$).

Subsequent data analysis confirmed and extended our earlier reporting of significant correlation between student learning outcomes (as represented by student final grade in the relevant course) and their use of engagement tools (discussions, mail) in fully online courses (Macfadyen & Dawson, 2010). Visualization of LMS use data for LMS-supported classroom-based courses again shows significant positive correlation between student participation in course-based discussions and their final grade (for number of discussion messages posted, $r = .83$, $p<.01$; for number of discussion messages read, $r = .95$, $p<.0001$, and for number of discussion replies posted, $r = .94$, $p<.0001$). A significant positive correlation with final grade is also observed with student use of LMS-based course content materials (0.89, $p<.001$), and also, most interestingly, with student visits to the “My Grades” tool (0.93, $p<.0001$) that allows students to monitor their own progress (Figure 4).

**Outcomes of participant observation**

The institution’s standing committee on learning technologies convened monthly during 2010 and 2011, with the goal of developing a vision, roadmap and plan for selection of the new institutional LMS. A detailed analytics report on the current state of implementation and use of the institution’s existing LMS was presented at an early stage in this process. Subsequently, meeting minutes and reports on later stages of decision-making were made available to the university community (data not shown). From review of these documents, and from participation in continuing committee discussions, we observed that although completion of the current state analysis was noted, no further references to or interpretations of the findings were made in later meetings or documentation.

**Discussion and implications**

**Benchmarking the institution’s LMS usage**

This e-learning analytics case study revealed multiple layers of data that can be re-purposed, aggregated and analyzed in new ways. By revealing details of actual LMS use patterns and their relationship to student learning outcomes, these data not only re-emphasize the value of the LMS in supporting student learning at the institution, but offer benchmarks by which the institution can measure its LMS integration both over time, and against comparable organizations.

The 2010 Campus Computing Survey (Campus Computing, 2010) reports that US public universities now make use of an institutional LMS in delivery of an average of 60% of their course sections, suggesting that LMS uptake in the university under study here, at only 21% of course sections, is comparatively low. Similarly, a 2007 study (Allen, Seaman, & Garrett, 2007) reported that US higher education institutions offered an average of 10.6% (median 5%) of their course sections in fully online mode, and that this number appears to be increasing rapidly as institutions further embrace distance and flexible models of education. In this case study, the institution’s small offering of online courses (1.6% of total course sections) again indicates a low level of penetration. Furthermore, at least 70% of teaching staff did not make any use of the institutional LMS during the 2009-2010 academic year.
While instructor use of the LMS may therefore be considered to be in the early adoption phase (Rogers, 1995), students are heavily exposed to the LMS, confirming the impression that the LMS is a core component of the institution’s overall learning experience. The vast majority (>80%) of students were enrolled in at least one course that made use of the LMS in 2009-2010, and 11% of students completed at least one fully online course. This latter figure, in particular, suggests that the student cohort is beginning to recognize the strategic advantages of online courses as they plan their timetables, meet program requirements and attempt to manage the time demands of work, study and commuting to campus. Workload and “time online” are core issues for teaching staff, and can also be significant areas of concern for students. With particular relevance to online courses, in which almost all course-related activity is mediated by the LMS, data show a very wide range of total student engagement time with peers and course materials across different disciplines. These data begin to provide lead indicators of the appropriateness of the course load as a result of the implemented learning activities.

A more detailed understanding of what, exactly, is occupying student time in LMS-supported course sites provides a more meaningful representation of how an LMS is being used, and therefore the degree to which LMS use complements effective pedagogical strategies. Contemporary educational theorists emphasize the importance of peer to peer interaction for facilitating the learning process (Astin, 1993; Chickering & Gamson, 1987; Light, 2001; Tinto, 1998). Nevertheless, when we examine the categories of activity that are occupying student time in LMS-supported course sites, it is clear that across the institution (and regardless of course mode), the dominant use of the LMS is for content delivery. This observation is further supported by the number of static text files contained within the LMS (Figure 2, text and pdf files). Adoption of technological innovations in a manner that simply replicates existing hegemonic practice (Reiser, 2007) is not limited to LMSs. Such “first stage” adoption appears to allow a familiarization phase, before broader innovations can be undertaken. It is only at this later innovation stage that learning technologies will be fully utilized to support a pedagogical practice of engagement that will significantly enhance the overall student learning experience. However, this will also necessitate the kind of cultural changes described by McWilliam (2005). A wealth of literature describes the enriched learning possibilities permitted by such a shift.

In relation to LMS functionality, while more than half of all LMS-supported courses at the case study university implemented a common suite of tools, mining the data on actual student use time for tools assists with overall interpretation for informed action. The current LMS tools that support online discussions, presentation and organization of course content, and assessment activities (usually quizzes) are the only tools which are heavily used. It can be argued that use of some LMS tools simply require little time investment – for example, tools that allows students to read a quick announcement, check their grades, or upload assignments. It is clear, however, that a range of available tools that could be used to increase student engagement and collaboration (MediaWiki, the Wimba suite of voice and video tools) remain poorly utilized. Further investigations are required to better understand why and how the adoption of these resources can be improved.

Together, these findings indicate that the institution has some distance to go in maximizing effective and strategic use of its enterprise LMS.

**Informing strategic planning?**

Although the data gathered in this case study analysis suggest that the potential use of the enterprise LMS is yet to be realized at this institution, it nevertheless confirms that the LMS is central to the student learning experience – a reality that should highlight the importance of careful planning for future learning technology uptake. The mandate of the standing committee on learning technology at this university is that it will support the institution's teaching and learning mission by assisting in the development of a campus-wide vision for technology use, and will lead the planning process for technology implementation. It is the only group explicitly tasked with integrating technology with the institution’s learning and teaching mission. With this in mind, it might seem surprising that subsequent steps in the institutional LMS review process did not appear to incorporate or build upon the intelligence revealed by this learning analytics exercise. While this committee might be considered to be the “powerful coalition” that Kotter (1996) identifies as a key actor in motivating and managing successful change, their subsequent discussions and deliberations did not include any critical consideration of current LMS use patterns in the development of a vision and strategic plan. Presentation of data indicating apparent correlations between student online engagement and student achievement did not catalyze debate about the pedagogical benefits of technology, or about whether the
institution as a whole appears to be making best use of available learning technologies. In essence, the findings derived from the learning analytics process failed to generate the sense of urgency or motivation for change as it related to technology adoption within the institution.

Diverse approaches to change management and leadership (see, for example, literature cited in Kavanagh & Ashkanasy (2006)) agree that development of an organizational vision, and a strategy by which to reach it, is a critical step. In this case study, learning analytics offered the institution a means of measuring its current state and future progress towards an institutional vision for teaching and learning with technology. However, through participant observation of committees responsible for moving the institutional LMS review and selection process forward, we noted that subsequent deliberations and decision-making focused almost exclusively on technical questions relating to “ease of migration.” Critical interpretation of the implications of data describing the institution’s current LMS use was almost entirely absent. These observations are reflected in public and private reports documenting the committee’s activities (not shown). While there is an obvious imperative to ensure that any new enterprise technology is functional, scalable and reliable, an exclusive focus on technology integration issues, in the absence of development of a pedagogical vision, quickly neutralizes the likelihood that learning analytics data may catalyze organizational change with a focus on the student experience and learning outcomes. A focus on technological issues merely generates “urgency” around technical systems and integration concerns, and fails to address the complexities and challenges of institutional culture and change.

The e-learning analytics data generated in this case study clearly demonstrate that some substantial changes are needed in order to better facilitate adoption and integration of learning technologies into daily curricular activities and support the ethos of student success to which the institution aspires. Recalling Gibbs’ (Gibbs, 2010) assertion, this institution already possesses the potential human, financial and technological resources (whichever new LMS it selects) to improve the quality of the education it offers. What will determine whether it succeeds or fails in this effort will be its ability to develop a clear vision for learning technologies and lead the cultural change that reaching it requires. Simple availability of new knowledge made available through e-learning analytics has, however, failed to influence institutional planning in this regard, and has failed to inform development of a strategic vision for learning technology at this institution. Interestingly, this mismatch between opportunity and implementation may be more widespread than enthusiastic analytics literature suggests. In their 2005 review of 380 institutions that had successfully implemented analytics, Goldstein & Katz (2005) note that analytics approaches have overwhelmingly been employed thus far “to identify students who are the strongest prospects for admission…[and]…to identify students who may be at risk academically” – that is, to improve enrollment and retention, rather than for institutional strategic planning. Similarly a recent survey of literature on implementation of educational data mining found that only a small minority of these report on the application of EDM to institutional planning processes (Romero & Ventura, 2010).

Why numbers are not enough

Why is it that the output of powerful learning analytics reporting processes, acknowledged by institutional leaders as giving new insight into organizational patterns and practices, fail to influence institutional planning and strategic decision-making processes? We suggest here that this may be the result of lack of attention to institutional culture within higher education, lack of understanding of the degree to which individuals and cultures resist innovation and change, and lack of understanding of approaches to motivating social and cultural change.

Although social systems such as educational institutions do evolve and change over time, they are inherently resistant to change and designed to neutralize the impact of attempts to bring about change (Kavanagh & Ashkanasy, 2006). This reality is reflected in Rogers’ theory of diffusion of innovation (1995), which attempts to model the factors that determine the adoption rate of (or, conversely, resistance to) new innovations. This model integrates variables at the level of the individual with variables introduced by the nature of the social system in question.

Perceived attributes of an innovation

Even if senior management and scattered individuals recognize the need for institutional change in order to better integrate technological innovations into teaching and learning, no vision or plan will emerge or be embraced without
the support of faculty and staff (Bates, 2000). Indeed, numerous writers have noted that a firm resistance to the changes that may be created by integration of e-learning must be expected (see Levy, 2003, and references therein). Rogers’ theory emphasizes the ways in which individuals will assess and resist proposed innovations according to the perceived attributes. Overwhelmingly, an individual’s reaction to change reflects their cognitive evaluation of the way in which a new event or context will affect their personal wellbeing (Lazarus & Folkman, 1984). When change is proposed, individuals will assess it situationally for its “relative advantage”: the degree to which change may offer something “better” than the current state. They will assess it for “compatibility”: the degree to which it is consistent with existing practice and values, and with needs of potential adopters. And they will assess it for “complexity”: the degree to which it is perceived to be difficult to understand or to use (Rogers, 1995).

Concerns surrounding academic workload have been commonly cited as reasons for a lack of adoption (Bates, 200; Levy, 2003; Macfadyen, 2004). For instance, faculty may view the introduction of technologies into teaching as a time-consuming imposition, as something that diverts them from current research and teaching activities, or as antithetical to the current institutional culture. Faculty and staff may see technology as bringing an extra (and unpaid) workload. Moreover, the potential for learning technologies to enhance teaching and learning may be poorly understood and incongruent with individual perceptions and beliefs surrounding good teaching practice. In particular, faculty may worry that spending time on technology will actually hamper their career due to poor evaluations of teaching. Such concerns are not without foundation: academic culture still rewards faculty for verifiable teaching expertise, publication output as a measure of research success, and independent achievement. The (often) context-specific nature of online teaching, the current lack of standardized methods of assessment of online teaching expertise, the time-commitment needed for quality instructional design, and the cooperative nature of effective team-based course development mean that incentives are often very low for faculty to invest time in working with technology (for overviews of these issues see Levy, 2003; Macfadyen, 2004; Oslington, 2005).

In institutions of higher education, senior representatives of university units—such as the Deans, Heads of Departments and other members of the senior administration participating in committees charged with LMS review and selection—are typically senior faculty members rather than professional managers. Rogers’ model illuminates for us that this cohort is most likely to evaluate proposed changes to the LMS infrastructure not by coherence with vision or strategy, but by assessing the degree to which any change will burden themselves and their colleagues with the need to learn how to use complex new tools, and/or the need to redesign their teaching habits and practices, without offering any appreciable advantage or reward. Information technology managers and staff similarly are most likely to assess proposals for new technology innovations from the perspective of workload and technical compatibility with existing systems, and have an even smaller investment in student learning outcomes. In this context, and in the absence of a strategic goal or vision (and of any clear incentives to strive towards such a strategic vision), analytic data reporting on current LMS data have little motivating power.

**The realities of university culture**

While faculty may be resistant to certain learning technologies, a more serious form of “institutional resistance” is found in the very culture of academic institutions—no less than a cultural clash. Bates (2000) characterizes the dominant Western university and college culture as a mixture of “industrial” and “agrarian.” In particular, the agrarian foundations of university culture is manifest today in a university structure in which learning is tightly regulated in a cohort/semester system, in which the faculty member is responsible for all aspects of teaching from selection of content to delivery to student assessment, and in which the accepted route for handing down knowledge is one of “apprenticeship” via supervised graduate study within a discipline (Macfadyen, 2004). In spite of the hierarchical management structures introduced by industrial models, the agrarian model gives insight into the persisting culture of cull faculty control of teaching.

At the institutional level, this “quality-and-effectiveness”-focussed culture offers a number of major obstacles to change: consensus governance (rather than industrial-style hierarchical management); faculty control over the major goal activities (teaching and research); an organizational culture that supports change by adding resources rather than by strategically reallocating resources, and a curriculum structure that makes false (though some would argue, necessary) assumptions about learner homogeneity (Volkwein, 1999). Change management theorists lay heavy emphasis on the role of leaders in motivating and managing successful change and innovation (Kavanagh & Ashkanasy, 2006), but while university presidents are expected to be inspiring leaders, any direct interference in
faculty democracy is not welcome. Similarly, introduction of policy that is seen to impinge on faculty autonomy in teaching is usually strenuously resisted, especially if it is perceived to derive from the “cost-consciousness-and-efficiency” culture of a management bureaucracy or corporate/industrial model for education (Macfadyen, 2004).

Where to from here?

Social marketing theorists (Kotler & Zaltman, 1971) and change management experts (Kavanagh & Ashkanasy, 2006; Kotter, 1996) agree that social and cultural change (that is, change in habits, practices and behaviours) is not brought about by simply giving people large volumes of logical data (Kotter & Cohen, 2002). These authors insist that in order to overcome individual and group resistance to innovation and change, planning processes must create conditions that allow participants to both think and feel positively about change—conditions that appeal to both the heart and the head. Learning analytics has the capacity to do both, but only if certain conditions are met.

Certainly, logical presentation of real institutional data can contribute to creating changes in thinking and behaviour, especially if it is used to highlight progress and room for growth against a backdrop of institutional targets and vision—and if participants are committed to the vision and motivated to achieve it. Interpretation remains critical. Data capture, collation and analysis mechanisms are becoming increasingly sophisticated, drawing on a diversity of student and faculty systems. Interpretation and meaning-making, however, are contingent upon a sound understanding of the specific institutional context. As the field of learning analytics continues to evolve we must be cognizant of the necessity for ensuring that any data analysis is overlaid with informed and contextualized interpretations.

In addition, we propose that greater attention is needed to the accessibility and presentation of analytics processes and findings so that learning analytics discoveries also have the capacity to surprise and compel, and thus motivate behavioural change. Rogers (1995) describes a further factor that influences resistance to innovation: “observability,” or the degree to which the results of change and innovation are visible to self and others. As Romero & Ventura (2010) note, to date, efforts to mine educational data have been hampered by the lack of data mining tools that are easy for non-experts to use; by poor integration of data mining tools with e-learning systems; and by a lack of standardization of data and models so that tools remain useful only for specific courses/frameworks. Collectively, these difficulties make analytics data difficult for non-specialists to generate (and generate in meaningful context), to visualize in compelling ways, or to understand, limiting their observability and decreasing their impact.

As governments and institutions further seek to establish quality measurements and demonstrate learning and teaching impact, learning analytics will be increasingly in demand. However, while learning analytics tools and processes will doubtless continue to rapidly evolve, research must also delve into the socio-technical sphere to ensure that learning analytics data are presented to those involved in strategic institutional planning in ways that have the power to motivate organizational adoption and cultural change.

References


Ley, S. (2003). Six factors to consider when planning online distance learning programs in higher education. *Online Journal of Distance Learning Administration, 6*(1).


Investigating the Development of Work-oriented Groups in an e-Learning Environment

Chia-Ping Yu and Feng-Yang Kuo

Department of Information Management, Tamkang University, Tamsui Hsien, Taiwan // Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan // cpyu@mail.im.tku.edu.tw // bkuo@mail.nsysu.edu.tw

*Corresponding author

ABSTRACT
In this study, we have investigated developmental patterns of virtual groups in the e-learning environment. Our findings suggest that for virtual groups formed for the purpose of e-learning, dependency and inclusion characterize the initial stage of group development, as such characteristics reinforce cooperative relationships and help to build a stronger social bond among group members. This is followed by the second stage, swift work, which enables participants to labor as a team and facilitates continual collaboration among members. However, the third stage, conflict, is inevitable, as conflicts provide important diagnostic evidence for each member to assess and adjust his or her values and preference. Finally, by overcoming conflicts, partners build strong bonding, which fosters intimate communication and provides many opportunities for frequent interactions that demonstrate concern and desire to satisfy the needs and wishes of one another. Our finding suggests that initial active dependency interactions in the first period provide a sense of coherence. Accordingly, the instructor of a virtual learning system should encourage learners to express their concern for one another in order quickly to build strong norms. In addition, as work intensifies, the instructor could aim to support both the effective interaction channels of groups as well as their task-focused activities. Finally, to improve members’ involvement and information sharing, the instructor could provide feedback to groups regarding their level of interactivity and encourage members to remind each other about the quality and quantity of individual contributions.

Keywords
Collaborative learning, Virtual group, Group development, E-learning

Introduction
Computer-mediated communication (CMC) technology such as the Internet has relaxed the limitations of proximity and structure on communication (Newhagen & Rafaeli, 1996; Hsu et al., 2008). It provides a unique opportunity for the learning and sharing of knowledge free from the constraints of time and place. One specific area of this application is Internet-based learning, often called e-learning, as many educators have begun to integrate the Internet into their work. The effectiveness of e-learning has been demonstrated by Dawson (2008), Hsu et al. (2008) and Rosson et al. (2009), who show that when the group is committed to a strong norm of cooperation, learners are able to share resources to improve their learning productivity in the group-based learning system. Other studies have also shown that collaboration in the CMC context can significantly improve learning interest and attitude (Vaughan, 2002) as well as facilitate better learning performance (Vaughan, 2002), increase cognitive capability, promote greater learning efficiency (Dawson, 2008; Ounnas et al., 2009), and enhance learning satisfaction (Rosson et al., 2009).

It is against such a background that the present research attempts to study how virtual groups develop and become actively engaged in collaboration in an e-learning environment. A virtual group is a collection of geographically distributed, functionally and/or culturally diverse entities that are linked by electronic forms of communication, the purpose of which is to make arrangements for alternative actions and to evaluate outcomes to determine future actions. In terms of e-learning, members coordinate with one another to set policies, make plans, execute these plans, and monitor their own activities to achieve the goal of completing the on-line course (Duphorne & Gunawardena, 2005). Research on physical groups suggests that such groups face many successive challenging phases (Wheelan & Mckeage, 1993; Wheelan et al., 1998). According to Wheelan et al. (1998), there are four stages of development for physical groups: dependency and inclusion, counter-dependency and fight, trust and structure, and work. Since there
has been little empirical research on the development of virtual groups, it is sensible to assume that the virtual group would go through similar stages of development (Duhorne & Gunawardena, 2005; Rosson et al., 2009). However, this assumption is open to challenge because virtual groups differ from the groups in the physical world in many ways. One important distinction is that relationships within the virtual context are tenuous, rendering virtual group membership much less distinct than physical-world membership groups (Järvenpää et al., 2004). For example, randomly allocating students to e-learning groups may lead to “orphan student” problems (Ounnas et al., 2009) and strong relationships among group members play a critical role in support and information exchange (Dawson, 2008). It has even been suggested that the development of the virtual group is likely to be ineffective because participants may withdraw from discussion easily (Ounnas et al., 2009). Therefore, the study of how virtual groups develop to become productive in learning is important to the designers of e-learning systems and educators who wish to employ them in their work.

**Group development**

Group development refers to the stages through which a group must successfully navigate in order both to structure relationships and roles within the group and to emerge as a mature, highly interactive unit capable of achieving its goals (Wheelan et al., 1998). For example, the study by Bales (1950) found that when faced with a decision, groups exhibited an increasing degree of maturity and performance over time that could be classified as linear. In his observation of task-oriented groups, Tuckman (1965) also identified four developmental phases: forming, storming, norming and performing, while Gersick and Hackman (1990) investigated student group development and proposed the punctuated equilibrium group model. Recently, Hsu et al. (2008) investigated the relationship between the norms of cooperation and resource sharing in e-learning group formation, while the study by Dawson (2008) suggested the need for virtual learning groups to succeed through social interactions to provide opportunities for knowledge construction and to generate a sense of group learning. In addition, Ounnans et al. (2009) considered the issue of group formation to propose a group-learning based system, while Rosson et al. (2009) investigated the role of the learning community for young women.

In brief, studies have shown that group effectiveness is linked to group development (Gersick & Hackman, 1990; Wheelan & McKeage, 1993). Summarizing previous studies (Tuckman, 1965; Gersick & Hackman, 1990), Wheelan (1994) suggested dividing group development into four stages: dependency and inclusion, counter-dependency and conflict, trust and structure, and work. Note that in this paper, we have chosen to use the term conflict to replace Wheelan's term fight.

**Stage one: Dependency and inclusion**

The first stage appears, on the surface, to be very harmonious (Wheelan & McKeage, 1993). Initially, these individuals often have only tentative relationships; however, what is lacking in terms of depth is compensated for in terms of breadth. Increased interactions among members assist group members to exchange both tangible and intangible resources, enabling the group as a whole to benefit. Members therefore look for guidelines that may alleviate their concerns and anxiety. There is also considerable imbalance with regard to relationship involvement, and members' optimistic assumptions about partners are not likely to be questioned. The main issues are related to psychological safety and inclusion, and members utilize a variety of strategies to gain the approval of the leader and other members. Consequently, while work occurs in this stage, members' mutual interactions are characterized not by task performance, but by dependency and inclusion, both of which facilitate the making of their social capital.

**Stage two: Counter-dependency and conflict**

This stage is characterized by conflicts among members, as well as between the leader and members. Conflict is somewhat of a two-edged sword. On the one hand, as the depth of interaction and interdependence increases, it is inevitable for people with different backgrounds gradually to discover each other's imperfections. On the other hand, if issues are avoided or if either partner withdraws from problem-solving interactions, the opportunity to use episodes of conflict to evaluate further a partner's abilities is likely to become entrenched in these dysfunctional patterns of conflict resolution (Peng et al., 2008). Thus, a partner's response during episodes of conflict may provide important
diagnostic evidence concerning the extent to which he or she is able to work to make the relationship successful. Ambivalence and an increase in the incidence of conflict often result from this enhanced awareness of the need for further compromise and adjustment. The motivation behind the struggle is to reduce anxiety by clarifying the goals and structure of the group. The extent to which conflicts can be successfully resolved is deemed to be essential to the development of cohesion and delineation of common values (Gersick & Hackman, 1990). This is demonstrated by good citizenship behaviors, such as helping others beyond one’s call of duty, and by more systematic organizational endeavors, such as task-relevant knowledge and tacit understandings to break boundaries.

Stage three: Trust and structure

Assuming that the conflict stage is successfully navigated, members of the group will feel secure with, and trusting of, one another and the leader. Not only do they feel no need to question their partner's motives and commitment, but their attention to structures and roles during this stage significantly increases the group's capacity to work effectively and productively (Newhagen & Rafaeli, 1996; Duphorne & Gunawardena, 2005). These embedded relationships involve the sharing of private, situated information and tacit knowledge, and relying on social, informal contracts between exchange partners (Meyerson et al., 1996). Participants are able now to exchange their comments openly. In addition, there is a shared sense of the expectations and standards of appropriate behavior, so that norms and rules of conduct can be decided upon.

Stage four: Work

For groups that successfully migrate through the previous stages, the next stage means that goals, structure, and norms are already established, and that the group can work more effectively. By this time, group members can receive and provide feedback about their effectiveness and productivity, and are able to take more responsibility than in the previous stages for jointly creating alternative directions and activities. Strong bonding exists because members now share similar beliefs, values and informality. This bonding, in turn, enables them to communicate regularly and interact socially to form more social capital (Wheelan et al., 1998; Duphorne & Gunawardena, 2005).

Methodology

Assessing group processes

In previous studies of physical groups, researchers have typically relied on videotapes or audiotapes, which are less intrusive than visible observers (Wheelan, 1994; Wheelan and McKeage (1993). In the CMC environment, CMC-technology can automatically record members' mutual interactions in a natural context. This method requires the collection of data from group interactions through CMC-technology and application of the content analysis approach to analyze the data generated. However, its limitations include difficulties in obtaining members' subjective perceptions as well as the historical and environmental contexts governing group development (Strauss & Corbin, 1998). It also may not record members' interactions that take place outside the CMC environment. Despite these disadvantages, collecting data from the discussions of virtual group members through the system log is the most viable method, since members of virtual groups, are separated in space and may interact asynchronously.

Participants

Observation of the class participating in the study took place over a sixteen-week period. Of the total of one hundred and fifty-eight participants, thirty-five were female. The majority of participants came from different areas. Their occupations were diverse, including teachers, programmers, managers, journalists, project leaders, doctors and company workers. They were assigned to twenty-five groups, the average group size consisting of 6 – 8 people, and were required to meet in the first week of the semester. One third of the participants had previous experience of courses offered by the same cyber-university. In the virtual classes, the students and the teacher kept on-line cyber-office hours every week. Students were evaluated on the basis of the quantity and quality of their contributions to discussions and final project reports. Each team had its own electronic BBS for collaborating with one another. In
addition, like the study by Järvenpää et al. (2004), each team held two F2F (Face-to-Face) meetings with the instructor, one in the middle and one at the end of the semester.

**Tasks**

Based on the task classification schema provided by McGrath and Hollingshead’s (1994), students in this study were asked to accomplish four types of tasks arranged sequentially: (1) Week 1 to Week 4, generate ideas; (2) Week 5 to Week 8, choose a preferred solution and set up the coordination mechanism; (3) Week 9 to Week 14, negotiate; and (4) Week 15 to Week 16, execute. The tasks are described below:

**Week 1–Week 4:**
Task Type: All teams were asked to determine the subject that they wished to pursue for the semester.

**Week 5–Week 8:**
Task Type: Each team was obliged to submit a project proposal, including details of each sub-task such as goal, tasks, procedure, assessment, and schedule.

**Week 9–Week 14:**
Task Type: Members were expected to search for and exchange information, clarify any task ambiguity, and monitor each other’s performance.

**Week 15–Week 16:**
Task Type: Execute. Teams were required to complete their respective projects and submit a written report showing their data analysis and collection methods, their findings and conclusions, which provided the important measures of the project’s quality.

**Content analysis**

Content analysis has the capacity to: (1) reflect patterns of group interactions; (2) reveal the focus of individual, group, institutional, or societal attention; and (3) disclose the relationship between intent and content (Krippendorff, 1980; Strauss & Corbin, 1998). For virtual teams, the members’ speech is a valuable resource in terms of describing members’ traits and states (Newhagen & Rafaeli, 1996) and is useful in content analysis to reveal individuals’ trust-related beliefs (Neuendorf, 2002).

The categories in our study are based on those developed by Wheelan (1994). Dependency sentences are those showing the inclination to conform to the dominant mood of the group, to follow suggestions made by the leader, and generally, to demonstrate a desire for direction from others. Tension-relief sentences are used to indicate avoidance of task and confrontation. Counter-dependency sentences assert independence and rejection of leadership and authority, or demonstrate the member’s attempt to lead. Conflict sentences utter criticism, argumentativeness, or aggression that conveys one’s struggle to overcome someone or something. Pairing sentences allow one to express warmth, friendship, support, or intimacy with others, while counter-pairing sentences are employed to indicate avoidance of intimacy and connection, and a desire to keep the discussion distant and intellectual. Finally, work sentences are those that represent purposeful, goal-directed activity and task-oriented effort. Note that tension-relief and conflict were referred to as flight and fight respectively in Wheelan’s work (1994).

**Procedure**

As in the approach adopted by Wheelan et al. (1998), the analysis unit in our study is the sentence. Because it is easier to agree on how to categorize one simple sentence than it is to agree on a whole conversation, in our content analysis, sentences are used as the analysis units to produce higher reliability among coders (Neuendorf, 2002; Krippendorff, 1980). A sentence may refer to one of seven coding categories. The discussion is coded by substantive comments, while redundant sentences are not coded.
In this study, two coders were employed and trained to ensure their skill and reliability in placing each unit into the appropriate Group Development category. The training data were collected from another course at the same cyber-university. A total of 12,567 postings were obtained for twenty-five groups across sixteen weeks. Several rounds of training practice were necessary, and the training was considered complete only when the reproducibility reliability of the results from the two coders exceeded 90% (Neuendorf, 2002; Krippendorff, 1980). Reproducibility reliability refers to the extent to which content classification produces the same results when the same text is coded by more than one coder.

To test the reliability of this study, we evaluated stability and reproducibility. In this study, each coder was asked to code a set of data at two different times. The degree of stability established by the two respective coders exceeded 90 per cent, satisfying the stability criteria (Krippendorff, 1980). Comparison of these two researchers’ coding results revealed that the reproducibility between the two coders was 94.37 percent, indicating an acceptable level of reliability (Neuendorf, 2002).

Furthermore, both face validity and semantic validity were employed in this study. To assess face validity in our study, the researchers were asked to step back in order to examine each concept objectively. Validity was considered to have been reached when all the researchers agreed on the data tapped into the desired concepts. For the classification to have semantic validity, coding units such as words classified together should possess similar connotations (Neuendorf, 2002). In this study, the two coders had two years’ experience participating in studying virtual teams and were familiar with the communication content. Semantic validity was achieved when the two coders examined the list of words placed in the same category and agreed that these words had similar meanings or connotations (Krippendorff, 1980).

**Results**

A total of 12,567 verbal sentences were obtained across sixteen weeks from twenty-five groups. Of these, 1,477 units (13.32% of the total) could not be classified into one of the seven categories of the integrated group development process because they were either duplicated or fragmentary. Therefore, the number of units used in the various analyses totaled 11,091.
To identify the characteristics of the four stages, ANOVA was performed to compare differences in verbal group developmental patterns. Significant differences were noted in dependency, tension-relief, and conflict and work (P-values were less than 0.001). The frequency of dependency was high in the first stage, as was tension-relief (Figures 1A and 1B). Conflict sentences peaked in the third stage, with work being high during the second and third stages (Figure 1C). The level of pairing increased during the final stage (Figure 1E). Note that Figures 1D and 1F show that very low numbers of counter-dependency and counter-pairing verbal sentences were generated in all stages, indicating a possible difference between virtual and physical groups.

**Dependency and inclusion**

Consistent with the study by Wheelan (1994), all of the virtual groups generated a large number of dependency and tension-relief sentences in the first stage (see Figures 1A and 1B). A typical example was, "No problem, we will follow the leader's commands ..." (in groups H2, L5, L17) or, "Our team members are experts ...... I believe we could accomplish our mission successfully.... This is great.... and I will follow it..." (in group H4). Dependency and tension-relief sentences may provide people with an alternative way to connect with others who commit to their interests or relational goals. Learners tended to be overly polite and tentative in an attempt to ward off potential group rejection, conforming to the leader to reduce anxiety levels. Thus, dependency and tension-relief communication reinforced the cooperative relationships, helping to build stronger social relations among the team members.

**Swift work**

In our data, the second stage was characterized by a surge of pairing and work-type sentences, although pairing sentences would surge again in the fourth stage. While this sudden surge in trust and work departs from the theory of integrated group development by Wheelan (1994), who predicts that members’ work will increase gradually over time, it is consistent with the finding of Wheelan and McKeage’s (1993) empirical study. Such a swift surge in work and trust may occur in both the physical and virtual teams (Meyerson et al., 1996; Järvenpää et al. 2004). With regard to groups formed for temporary purposes and under pressure to work, members emphasize speed and
confirmation information initially (Meyerson et al., 1996). As such, a person’s work-type sentences may serve to provide assurance to the group, therefore, permitting the upholding of the individual’s reputation.

Furthermore, this role-based interaction allows participants to cooperate with one another without spending time building relationships. In CMC environments, the lack of social cues and individual features may enhance, rather than diminish, the salience of stereotypical group features so that people may perceive others not as individuals with a range of idiosyncratic characteristics, but as representatives of social groups (Spears & Lea, 1992; Postmes et al., 1998). Accordingly, in CMC environments, members of temporary groups may strengthen the cognitive process of social categorization to form stereotypical impressions of other members (Spears et al., 2002). The categories affect expectations of good will or ill will, and encourage swift work and swift trust. For example, labels such as “leader” or “active participant” are quickly confirmed by their sentences. In group H7, Jack volunteered to become the leader. In order to create a brave leader’s image, he posted large amounts of work-related information such as, “I’ve got several interesting reports in XX magazine… Hope they are helpful for our project…..” In another group, H5, the leader actively encouraged Mary to engage in their teamwork, even though she had no cyber-university experience. To be a “good” member, Mary always posted her class notes, homework, technical reports, and interesting URL on their group board and said, “If you need more information, please let me know…..” Thus, Järvenpää et al. (2004) assert that members of virtual groups rely heavily on category information at the beginning of group development and act accordingly.

It should be noted that the work sentences may not be helpful in completing the group project in the early stage, as illustrated by the statement made by the leader of H7. However, this should not be considered negative. Swift work-type sentences allow participants to work as a team and promote continual collaboration. They are relied on by members to establish their position in the group. Even though members are not aware of the covert purpose of a discussion in the initial stages of group development, they generally respond to others’ demands in order to be viewed as influential. In other words, work-type sentences at this stage are frequently made by individuals to indicate to other members their willingness to collaborate, leading to the possible establishment of a cooperative and responsible group.

**Conflict**

While swift work creates an active and effective virtual group image that allows members to collaborate quickly, it runs the risk of the group losing touch with reality. When words are said but results are not delivered, members are unable to accomplish their mission and become frustrated, which inevitably leads to conflicts. Thus, our study shows that in the third stage of virtual group development, the number of conflict sentences peaks (see Figure 1C).

According to Wheelan (1994), conflict-type sentences pave the way for the resolution of conflicting needs and preference differences, allowing the partners to further consolidate their state of trust. That is, when conflicts emerge and are resolved successfully, members’ relationships may be permanently stabilized (Meyerson et al., 1996; Järvenpää et al., 2004). H5 is an example of successful conflict resolution, where members expressed doubts, such as, “It is not a good idea… we could try another method….“or “You need to revise…..” Yet, at the same time, they provide warmth and support to encourage their partners, as revealed in expressions such as, “You did a good job…..” “We could overcome it together…..” or “I can provide you with some relevant information about…..” After several weeks, H5 team members collaborated to work out the project schedule together and expressed a sense of satisfaction in on-line discussions. By that time, they had already agreed to a common goal and shared a structured communication process. One of them stated, “I enjoyed very much working with you. This is an impressive negotiation process…..” and another said, “I think it is a great experience that we could build our ideas upon this project…..” In H5 group, learners were more likely to become involved in vicarious learning through comparison of their own attainments with those of their partners. Ultimately, H5 members dealt successfully with conflicts.

However, conflicts may also lead to dysfunctional groups, partly because members fail to act as expected, and partly because they prompt individuals to question their initial judgment concerning the categories obtained earlier. L3 is an example of failure. From the second stage to the third, L3 members disagreed with their final report draft and became embroiled in serious conflict. The leader commented, “You should be more responsible! …Look at what you have handed in!” One member did not accept the leader’s criticism, saying, “It is unfair…..I have done everything I could…..” Other members also failed to act as expected, and negative emotions prevailed. Ultimately, the L3 group failed
in their final project. Members felt frustrated and regarded their group as unreliable, which led to a breakdown in collaboration.

To sum up, conflict-type sentences provide important diagnostic evidence to each member involved in virtual group development to assess and adjust his or her values, needs, and preferences in order to make the successful transition into the next stage. If virtual groups avoided confrontation or withdrew from problem solving, they would lose opportunities to evaluate problems from within the group.

Trust

The percentage of pairing sentences peaked in the fourth stage (Figure 1E). To repair the wounds caused by conflicts, partners had to demonstrate concern and desire to satisfy others’ personal needs, preferences, and wishes. Consequently, the team’s social capital was strengthened, and members became engaged in a shared practice. The leader of H7, for example, invited all members to resolve problems together when conflicts arose. They used expressions such as, “This idea is great and we need to talk about it together…” or “We would like to know everyone’s comments…” In essence, partners truly felt they were accepted for who they were, and developed mutual understanding. In such a situation, participants’ expectations accorded with their general feeling of solidarity involvement. Moreover, for successful group development, a sense of trust and group identity could grow on the basis of the assessment of the feelings and attitudes of the partners toward one another. For example, in group H2, members in the third stage had conflicts, illustrated in expressions such as, “Our project is not qualified…” However, in the same stage, they also used phrases such as, “I feel proud of our group, and I know you work very hard…” or “We could overcome it together…” These exchanges led H2 members to believe that they could accomplish the group goal in spite of their conflicts.

By this time, with respect to high interactivity groups, their social capital is clearly developed. Let us consider H2, for example: as a group, strong trust, shared knowledge, and mutual obligation develop in H2, and learners voluntarily collaborate in data collection, data analysis, or assist other members. However, in dysfunctional groups (L09 and L11), members either put up strong resistance to the suggestions of others or withdraw from the problem-solving activity.

In short, virtual groups produce dependency, tension-relief, work, conflict, and pairing sentences in all stages, although the number of counter-dependency and counter-pairing sentences are few. In our study, dependency and tension-relief sentences characterize the first stage, swift work the second, conflict the third, with the pairing verbal category characterizing the final stage.

Discussion

In this section, we will compare the results of our study with those of Wheelan and McKeage’s (1993) study. Figures 2 through 4 compare the patterns of the verbal category in the two studies. The first similarity between these two is found in the dependency-type sentences (see Figures 2A and 2B). It appears that in order to build group relations and a sense of involvement, both FtF and virtual groups display a large number of dependency sentence to reinforce collaborative relations among the team members. Second, there is a similar pattern of pairing-type sentences in FtF and virtual groups (see Figures 3A and 3B). In the pairing conversations, members act as supporters, attempting to get their work back on track and sharing different opinions. They serve to smooth out the anxiety and embarrassment aroused by conflict sentences. Third, both virtual and FtF groups generate a considerable number of work sentences during the early stages of group development (Figures 4A and 4B). In order to secure their position in the early stage of group development, members rely on category information as the basis of mutual interaction, and seek to listen, observe, practice, and ask for feedback from work-type sentences. These pave the way for later conflicts and their resolution.

It is worth pursuing further analysis of the work statements in our study from the perspective of interactivity, which refers to the extent to which “in a given series of communication exchanges, any third or later transmission is related to the degree to which previous transmissions” (Rafaeli, 1988, p. 111). In our study, while there was a total 5,823 work-type statements, there were only 1,477 statements replying to previous postings. In other words, only around
25.36 per cent of work conversations could be shared and discussed with each other. Although it is not clear if it also appears in Wheelan and McKeage’s study (1993), we suspect that one reason for our study’s low interactivity is the existence of free-riders, i.e., virtual group members who do not respond to others and do not participate in group work. According to Spears et al. (2002), in CMC environments, depersonalization may cause people to be less accountable to others due to the lack of identity, leading to negligence of important information or withdrawal from discussion. This effect of depersonalization can be seen in Figures 5A, which shows a decreasing pattern of tension-relief statements in our study, and Figure 5B, which shows that the number of tension-relief statements remains high over time in Wheelan and McKeage’s (1993) study. Thus, it is possible for CMC to become a space where members deal with feelings of anxiety and embarrassment by remaining silent and, subsequently, may not feel the need to employ tension-relief sentences to evade confrontation. For example, when members of groups T09 and T25 had problems with their work, they seemed to ignore their mistakes. Our data show that they never addressed the first failure and, worse yet, their level of communication dwindled substantially afterward. As others failed to act as expected, many members remained silent to avoid negative emotion. Eventually, members seldom verbalized their commitment, excitement, and optimism. Consequently, silence can become a double-edge sword in the CMC-world: while it helps to hide negative feelings, silence also allows negative feelings to build up.

Figure 2A. The pattern of dependency-type sentences in our study

Figure 2B. The pattern of dependency-type sentences in Wheelan and McKeage’s study (1993)

Figure 3A. The pattern of pairing-type sentences in our study

Figure 3B. The pattern of pairing-type sentences in Wheelan and McKeage’s study (1993)

Figure 4A. The pattern of work-type sentences in our study

Figure 4B. The pattern of work-type sentences in Wheelan and McKeage’s study (1993)

In addition, the pattern of conflict sentences for the virtual groups differs from that of the FtF groups (Figures 6B and 6A), with the percentage of conflict sentences for the former group being higher than for the latter. In the virtual world, lack of authority and loose membership may provoke a greater exchange of extreme arguments (Spears & Lea, 1992). In other words, members of the virtual groups may use conflict-type sentences more frequently than FtF groups. According to Gunawardena's study (1995), the removal of social cues, particularly indicators of status,
power and leadership) eradicates those social inhibitions or normative constraints that could act as a brake on the generation or articulation of extreme arguments. Thus, participants in virtual groups might use conflict statements more frequently than FtF groups because virtual group members are not as concerned with maintaining satisfactory personal relationships as in those involved in FtF communication.

Finally, while Wheelan and McKeage’s study shows that FtF groups produce a large number of counter-dependency and counter-pairing sentences, groups in our study rarely employ these two types (see Figures 7A, 8A, 7B, and 8B).
This could be due to the difference between FtF communication and CMC. According to Wheelan (1994), when group members feel hurt by authority, they become hostile or aggressive, resorting to counter-dependency sentences to retaliate. However, in CMC communication virtual groups, members can easily withdraw from the virtual space and cut themselves off from other members. Similarly, there is probably little need for members of virtual groups to rely on counter-pairing sentences to indicate the avoidance of intimacy and connection.

Conclusion

In this study, we have investigated development patterns of virtual groups in the e-learning environment and have compared the differences between virtual and FtF groups. Our findings suggest that virtual groups produce tension-relief, work, conflict, and pairing sentences in all stages; however, their counter-dependency and counter-pairing sentences are few. In addition, we have found dependency and tension-relief sentences to characterize the first stage; swift work, the second; conflict, the third; and trust, the final stage. Dependency and tension-relief communication reinforce cooperative relationships and then help to build stronger social relationships among the team members. Swift work-type sentences allow participants to collaborate as a team and facilitate continual cooperation among members. Gradually, conflicts provide important diagnostic evidence for each member to assess and adjust his or her values, needs, and preferences in order to make a successful transition into the next stage. Finally, overcoming conflicts, partners use pairing sentences to demonstrate concern for and desire to satisfy others’ personal needs, preferences, and wishes.

Moreover, comparing our study with that of Wheelan and McKeage (1993) reveals that virtual and FtF groups have similar patterns in terms of dependency, pairing, and work-type sentences, but differ with respect to the tension-relief and conflict statements, due probably to depersonalization. It is likely that there is little need for members to rely on counter-dependency and counter-pairing sentences to indicate the avoidance of connection and intimacy, because they are able easily to withdraw from the virtual space and cut themselves off from other members.

This study provides a basis for theorization about virtual group development. With findings that are consistent with those of Dawson (2008), who suggested the need for virtual learning groups to succeed through social interactions to provide opportunities for knowledge construction, our study is useful in developing guidelines for managing virtual groups in education settings. First, initial active dependency and tension-relief interactions in the first period provide a sense of coherence. Accordingly, instructors on virtual learning systems should encourage learners to express their concern for one another in order quickly to establish strong norms (Hsu et al., 2008). To reduce the problem of low interactivity caused by free rider effect, it is also necessary for the instructor to make clear, as early as possible, that group interactions will be monitored constantly and used for assessing group performance (Roberts & McInnerney, 2007; Peng et al., 2008). Allowing comprehensive tracking of both the individual’s and the group’s on-line activity is truly an advantage of the CMC environment. The instructor can employ a marking scheme that provides different marks to group members based upon their individual contributions to group interactivity and build into the assessment process an element of peer and/or self-assessment (Roberts & McInnerney, 2007). This may reduce social loafing by enabling equitable allocation of outcomes to individual input.

Next, as work intensifies, instructors could aim to support both the effective interaction channels of groups as well as their task-focused activities. For example, the virtual learning system might provide clear learning-group goals and milestones so that members can easily organize their work to meet project requirements. This would reduce the likelihood of learners’ commitment to work becoming nothing more than empty promises. The instructor could also provide feedback to groups regarding their level of interactivity and encourage members of respective groups to remind each other about the quality and quantity of individual contribution (Roberts & McInnerney, 2007; Peng et al., 2008).

Third, instructors of virtual groups should carefully consider issues related to conflict resolution to ensure the viability of these groups. For example, instructors and learners should be discouraged from making assertions about the personalities and motives of others or about others’ incapacities to deal effectively with problems. Finally, to improve members’ involvement and information sharing, the instructor should carefully consider issues for virtual teamwork conflict resolution such as dependability, conflicts, and trust in order to ensure the viability of e-learning groups. Such comments typically cause the group to experience doubt and cynicism about its problem-solving and decision-making effectiveness.
An important limitation of our study is that while our analysis has relied solely on discussions conducted on the Internet, these discussions do not include all of the communications within each team because the members maintained contact both through the Internet and through F2F meetings. These groups were work-oriented virtual groups but their communications did not happen in a pure CMC-environment. In addition, as the analysis is based on data collected from a single class in one specific school in a CMC-environment (C course in the cyber-university), the results may not be generalizable to other virtual contexts. Although we observed naturally occurring groups, these groups were comprised of students whose rewards were grade-based. The group size in our study was around 6-7 persons. In this situation, one person or small group could conceivably have completed the project although the workload would have been immense. Thus, because this study was conducted within a single cyber-class, generalization of the results to other virtual groups may be limited. Finally, there were several differences between Wheelan’s (1993) sample and ours. Group members in Wheelan’s study were training workshop participants, and the group life was one week. In contrast, our samples consisted of learners at cyber-university, who were members of groups with a life of sixteen weeks. However, there were also points of similarity between Wheelan’s sample and ours; for example, both samples were learning groups. Despite the above-mentioned limitations, the results of this study may suggest interesting implications for the study of virtual group development.

References


Alignment of Teacher and Student Perceptions on the Continued use of Business Simulation Games

Yu-Hui Tao¹, Chieh-Jen Cheng² and Szu-Yuan Sun²
¹Department of Information Management, National University of Kaohsiung, Kaohsiung, Taiwan // ²Department of Information Management, National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan // ytao@nuk.edu.tw // u9524819@ccms.nkfust.edu.tw // sunnyy@ccms.nkfust.edu.tw

(Submitted April 21, 2010; Revised December 22, 2010; Accepted March 6, 2011)

ABSTRACT
The higher education system in Taiwan has increasingly adopted business simulation games (BSGs) in recent years. Previous BSG benefit research has shifted focus from learning performance to motivation due to mixed results. One recent study empirically investigated student perceptions on the continued use of BSGs; however, the counterpart of higher education teachers’ perspective was absent. This study confirms the student perception model from the teachers’ viewpoint via a comparable research design. Empirical evidence shows a significant perceptual gap between teachers and students, implying that adopting BSGs on the part of teachers may be challenged or impaired. Detailed suggestions on gap alignment are discussed, including how teachers may take advantage of students’ positive perceptions. Among the theories used in the student perception model, agency theory may deserve further research, since incentives are typically considered a useful means of facilitating general teacher-student dynamics. However, all related hypotheses were unsustained in this study.

Keywords
Business simulation games, Agency theory, Expectation confirmation theory, Technology acceptance model

Introduction

Most computer games emphasize commercial benefits or leisure interests, while only a small percentage of these simulation games are applicable to the business management curriculum in higher education (Virvou & Katsionis, 2008). However, over 95 percent of AACSB accredited business schools in the U.S.A. have incorporated business simulation games (BSGs) in their curricula, since over 200 games were available in the market in the 1980’s (Wellington & Faria 1995). This high percentage of BSG usage in higher education has attracted academic research on BSG learning performance. Extent research on BSG has presented mixed results between learning performance and attitude (Vaidyanathan & Rochford, 1998) and thus the focus has shifted to other non-performance perspectives, such as students’ motivational effects (Schwabe & Göth, 2005). Since designing a quality BSG may not be easy and cost-effective, Amory (2007) proposed a theoretical framework, the Goal Object Model (GOM) II, with comprehensive perspectives of game, visualization, elements, problems, and social spaces for educational game development, which can also serve as a mechanism to evaluate the use of computer games in the classroom. In contrast, Prensky (2008) and Lim (2008) summarized an alternative approach of how student users can act as designers to develop useful games to support learning.

In theory, as an educational technology, a BSG must attain certain learning values, provided that its applicable scope is identified. Evident examples range from a general learning platform for facilitating e-learning (Tao, 2008) to a student response system for increasing classroom interactions (Kay & LeSage, 2009), or to speech-to-text recognition (STR) technology for assisting non-native speakers and learners with hearing impairment and learning difficulties (Wald, 2004). Despite the potential benefits of educational technologies, issues on its adoption among teachers and students hinder its spread in actual practice. As implied in GOM II by Amory (2007), the root causes may be that BSGs are extremely expensive and time-consuming to develop or low cost-effectiveness for highly customized needs, which transmit to adoption issues when they are finally available in the market. Among many studies of users’ attitude toward BSG, only one study focuses on students’ continuance usage in higher education (Tao, Cheng, & Sun, 2009). However, it merely presents students’ perceptions on the continued use of BSGs without considering teachers’ confirmation.

The importance of the teacher-student dynamic regarding the adoption of educational technology is evident in the literature. For instance, e-learning strategies suggested in a teacher study (Tao & Yeh, 2008) were further refined when the comparative study on students was supplemented (Tao, 2008). The study showed that stakeholder perceptions are important in presenting the overall perspective of issues concerning common interests in the teacher-
student dynamic. Ideas surrounding digital natives and digital immigrants (Prensky, 2001) may represent the current profiles of students and teachers regarding the adoption of any educational technology. The digital disconnect shows the gap between Internet-savvy students and their teachers (Levin & Arafeh, 2002) such that “the gap between students’ perception of technology and that of the faculty continues to widen” (The New Media Consortium and the EDUCAUSE Learning Initiative, 2008). This realization implies that students and teachers may have the same preferences for learning a new technology, but the context in which these preferences play out may vary.

Although Taiwan initiated BSGs as early as 1973 (Kung-Hwa Management Foundation, 2010), only in recent years have higher education institutes adopted BSGs for classroom activities and conducted national competitions in this area. The study on students’ continued use of BSGs in Taiwan’s higher education institutes (Tao et al., 2009) fits the potential issues of the teacher-student dynamic, i.e., the lack of teacher perception to comprehend the perspectives of continued use of BSGs.

To fill in the missing aspect in the study of Tao et al. (2009), this study addresses understanding and aligning teacher perceptions on the continued use of BSGs with those of students. The current study focuses on identifying the discrepancy in the perception of teachers and students on the continued use of BSGs. This may serve as a valuable reference for teachers in evaluating the significance of technology, as well as improving their instructional methods and skills using BSGs. The latter part of this article presents the research model and study results of Tao et al. (2009); the consequence of misalignment; a brief introduction of the research method and design; the comparative data analyses, discussion, and implications; and finally, the conclusions based on research results.

Background

This section provides a summary of the student study in Section 2.1, followed by the consequence of misalignment studies in Section 2.2.

Student model and results

Figure 1 shows the integrated research model for students. The joint Technology Acceptance Model (TAM) and Expectation Confirmation Theory (ECT) serves as the core foundation, enhanced by integrating the Agency Theory (AT), emotional factors, and learning factors to form the overall research model.

![Figure 1. Student’s continuance use of business simulation games model](image)
Figure 1 presents hypotheses H₁ to H₁₇, summarized in Table 1 based on the original theories with supporting references. Measurement items for factors in the hypotheses can be modified and adopted from the supporting references. When these measurement items were implemented, the context of BSG in class was mentioned in relation to their answers.

**Table 1. Hypotheses and supporting references**

<table>
<thead>
<tr>
<th>Expectation Confirmation Theory related hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₁. Satisfaction has a noticeable impact on the intention of user’s continuous use. (Bhattacharjee, 2001a; Roca et al., 2006; Chiu, Hsu, Sun, Lin and Sun, 2005)</td>
</tr>
<tr>
<td>H₂. The confirmation degree of the students using the business simulation games has a noticeable impact on their level of satisfaction. (Thong et al., 2006; Roca et al., 2006)</td>
</tr>
<tr>
<td>H₃. The student’s learning performance has a noticeable impact on their level of satisfaction. (Oliver and DeSarbo, 1998; Tse and Wilton, 1998)</td>
</tr>
<tr>
<td>H₄. The student’s learning performance has a noticeable impact on their level of confirmation. (Churchill and Surpreman, 1982; Tse and Wilton, 1998)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Agency Theory related hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₅. Incentives have a significant impact on the learning performance. (Bhattacharjee, 2001b; Eisenhardt, 1989)</td>
</tr>
<tr>
<td>H₆. Incentives may receive the interference of goal conflicts because of the learning performance. (Bhattacharjee, 1998)</td>
</tr>
<tr>
<td>H₇. Incentives may receive the interference of risk-aversions because of the learning performance. (Bhattacharjee, 1998)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Technology Acceptance Model related hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₈. Perceived usefulness has a significant impact on the satisfaction level. (Bhattacharjee, 2001a; 2001b; Devaraj et al., 2002)</td>
</tr>
<tr>
<td>H₉. Perceived ease of use has a significant impact on the satisfaction level. (Devaraj et al., 2002; Roca et al., 2006; Thong et al., 2006)</td>
</tr>
<tr>
<td>H₁₀. Perceived usefulness has a significant impact on the learning performance. (DeLone and McLean, 2003; Doll and Torkzadeh, 1998)</td>
</tr>
<tr>
<td>H₁₁. Perceived ease of use has a significant impact on the cognitive playfulness. (Venkatesh, 2000; Liao et al., 2007; Heijden, 2003)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning Factors related hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₁₂. The learning motivation of a student has a significant impact on his learning performance. (Pintrich &amp; De Groot, 1990)</td>
</tr>
<tr>
<td>H₁₃. The classroom atmosphere has a significant impact on the learning performance. (Moos, 1971; Deng, 1992)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emotion Factors related hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₁₄. Cognitive playfulness has a significant impact on the satisfaction level. (Webster et al., 1993; Moon and Kim, 2001)</td>
</tr>
<tr>
<td>H₁₅. Perceived attractiveness has a significant impact on the perceived usefulness. (Heijden, 2003)</td>
</tr>
<tr>
<td>H₁₆. Perceived attractiveness has a significant impact on the perceived ease of use. (Heijden, 2003; Tractinsky et al., 2000)</td>
</tr>
<tr>
<td>H₁₇. Perceived attractiveness has a significant impact on the cognitive playfulness. (Heijden, 2003)</td>
</tr>
</tbody>
</table>

Figure 1 shows the results of model verification. Most of the ECT hypotheses and all hypotheses regarding the emotional factors have been established. In contrast, none of the AT hypotheses and hypotheses on learning factors have been established. In addition, all hypotheses of TAM have been established, except that on satisfaction.

Playfulness in BSG games has an important influence on students’ continued use of these games, while its perceived usefulness has an important influence on perceived playfulness. The attractive features of the interface influence student perceptions of usefulness, ease of use, and playfulness of business games.

**Consequence of misalignment**

The most recognized misalignment study may be the PZB service quality model (Parasuraman, Zeitham, & Berry, 1985), in which five different perception gaps resulted in different consequences of poor service quality issues.
Similarly, Rovai, Ponton, Derrick, & Davis (2006) reported that an online course delivery medium introduced misalignment to some students between their preferred learning environment and actual learning environment. In general, educational technology as an aid within the teacher-student dynamic poses similar gap issues on the perception of teachers and students on the use of technology.

Similar to any information system implementation, the consequence of BSG adoption could be challenged or impaired if not successful (The Standish Group, 1995). Without efficient developments on the initial use of an educational technology, teachers would not be able to develop online teaching methods or go beyond the basic transmission model, considering the increased teaching workload (Palmer & Holt, 2009). Even worse, teachers might stop using educational technology, or even choose to not use it in the first place, if the adoption is voluntary. Students typically hold a more positive attitude towards the use of educational technology compared with teachers or members of the administrative staff (McGill & Hobbs, 2008; Palmer & Holt; 2009). Misalignment may lead teachers to use inappropriate teaching or instructional methods in classes, and thus create a bad BSG experience for students.

Many educational settings manifest the consequences of misalignment in related educational technologies, as in the cases of previously mentioned e-learning (Tao & Yeh, 2008; Tao, 2008), interactive response systems (Kay & LeSage, 2009), and STR technology (Wald, 2005) approaches.

Research Methods

To conduct a comparative study, this extended study adopted the same methodology used in Tao et al. (2009) for measuring student perceptions. This methodology uses the integrated research model in Figure 1, the hypotheses in Table 1, and the 52 questionnaire measurement items. Except for their expressions, the measurement items were adjusted so that teachers could appraise how students think about using BSGs or to provide information on how they think. Either way provides perspectives to teachers to gauge the gap of student perspectives using the same measurement method. Each question was measured with a seven-point Likert-type scale ranging from 1 (extremely disagree) to 7 (extremely agree).

This survey questionnaire was used to collect data, including the 52 measurement items and some personal data items, including gender, age, location of affiliation, rank of professorship, habit of playing computer games, proportion of class time using computer simulation games, and factors for the continued use of BSGs. The questionnaires were sent to college teachers in Taiwan’s higher education institutions who have adopted BSGs for their class activities. Statistical methods of descriptive analysis and Structural Equation Modeling (SEM) were used to analyze the responses of teachers on the same set of hypotheses intended for students as described in Tao et al. (2009). Similarities and differences were also analyzed and interpreted.

Data Analysis

Descriptive analysis of means and Partial Least Square (PLS) of the SEM software used in the data analyses are shown below.

Sample profile

Data from teachers were collected during a three-week period starting from the second week of June 2009, and 58 valid questionnaires were returned from the 122 higher education teachers initially identified to be BSG adopters. Before describing the teachers’ general profiles, the students’ profiles in Tao et al. (2009) were summarized for comparative purpose: 34.1 percent of students were male and 47.6 percent developed the habit of playing computer games. Most students belonged to universities in Southern Taiwan (58.4 percent), and majored in management (97.3 percent). Among the surveyed students, 76.2 percent attended only one class, which used BSGs, and the BOSS (88.1 percent) and Chain Store Master (34.1 percent) were mostly used in their classes. Statistics indicated that 56.2 percent of students used BSGs for an hour or less after their classes. The three major factors influencing students to use BSGs in their classes were: (i) interesting content of the game (63.2 percent), (ii) it must tally with the basic
requirements of the course (56.2 percent), and (iii) it can increase course participation (56.2 percent).

The teachers consisted of 62.1 percent male and 37.9 percent female. The highest percentage of age group ranged from 36 to 45, and totaled 63.8 percent. The survey was filled by 41.4 percent of the northern schools and 36.2 percent of the southern schools; of these, 96.6 percent were management schools. Assistant professors constituted 37.9 percent of the survey responses and 31.0 percent from associate professors. Most teachers who answered the survey questionnaire did not habitually operate computer games (86.2 percent). In addition, 87.9 percent of teachers experienced using the BOSS game, and only 24.1 percent experienced the Chain Store Master game. Teachers who used BSGs in one class made up the highest percentage, i.e., 48.3 percent. Teachers using BSGs in their classes not exceeding the time proportion of 1/6 (32.8 percent) exhibited a higher percentage compared with those using BSGs in classes not exceeding the time proportion of 1/3 (24.1 percent). The three factors influencing teachers’ continued usage of BSGs are as follows: (i) it must tally with the basic requirements of the course (82.8 percent), (ii) it can increase course participation (81.0 percent), and (iii) it can improve learning performance. These factors generally match the teaching objectives of teachers for using business games (Faria & Wellington, 2004).

This research included several interesting observations. First, the major difference in terms of motivation for teachers and students to continue using BSGs based on the three factors is that students felt it was interesting, while teachers felt that it might improve student-learning performance. Second, there was a disconnection between teachers and students in terms of personal computer-game playing habits. While 47.6 percent of students played computer games, 86.2 percent of teachers did not. The 35 percent difference strongly reflects the anecdotal evidence that students are more comfortable with technology as they are generally younger than teachers (McGill & Hobbs, 2008). Since over half of the sample students did not play computer games, the digital disconnect is obvious, but not as significant as the relationship of digital natives and digital immigrants as stated by Prensky (2001). Third, the time spent in a BSG was not as significant as we believed. Most teachers used BSGs for less than one-third of each class, while 56.2 percent of students used BSGs for less than an hour after class. This fits the factor influencing both teachers’ and students’ continued use of BSGs, i.e., “It must tally with the basic requirements of the course.” In other words, BSGs were used as teaching aids for certain management subjects or for serving a class activity when appropriate. Fourth, the survey forms were completed by 41.4 percent of teachers in the North and 58.4 percent of students in the South. This is reasonable, because more teachers in the North have adopted BSGs based on our compiled list of 122 teachers. As for the students, Tao et al. (2009) identified more heavy-usage teachers in the South to enable their students from multiple classes to accomplish the survey forms. The teacher sample represents well its population, while the student sample could not be determined since no estimate is available.

Reliability and validity

The factor loading in each item is well above 0.5, and all the internal consistency composite reliability (CR) values exceed 0.7, which accords with the study of Hair et al. (1998). The data demonstrate overall good reliability.

All AVE values lie between 0.64 and 0.88, satisfying the > 0.5 (Forrell & Larker 1981). This indicates that the convergent validity of this research is acceptable. The discriminant validity indicates that every aspect of the square root of AVE exceeds the construct and other related coefficients of the construct, thereby tallying with the suggestions of Chin (1998) stating that discrimination between the constructs exists.

Comparative descriptive analysis

As shown in Table 2, the mean values all lie between 4.34 and 5.05 for students and between 4.56 and 5.54 for teachers. Statistically speaking, we confirm significant perceptual gaps between teachers and students on perceived usability, playfulness, classroom climate, incentives, and learning performance. Visually speaking, the mean values calculated for teachers are generally higher than the ones for students, against the impression that students generally hold a more positive attitude towards educational technology (McGill & Hobbs, 2008; Palmer & Holt, 2009).
<table>
<thead>
<tr>
<th>Constructs</th>
<th>Items</th>
<th>Mean (Students/Teachers)</th>
<th>Noticeable difference between the teachers and the students (t test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness</td>
<td>PU1</td>
<td>4.95 /5.29</td>
<td>4.99/5.47</td>
</tr>
<tr>
<td></td>
<td>PU2</td>
<td>5.04 /5.72</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>PU3</td>
<td>4.99 /5.40</td>
<td>*</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>PEU1</td>
<td>4.65 /4.86</td>
<td>4.69/ 4.86</td>
</tr>
<tr>
<td></td>
<td>PEU2</td>
<td>4.68 /4.91</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>PEU3</td>
<td>4.75 /4.83</td>
<td>*</td>
</tr>
<tr>
<td>Perceive attractiveness</td>
<td>PA1</td>
<td>4.31 /4.57</td>
<td>4.49/ 4.68</td>
</tr>
<tr>
<td></td>
<td>PA2</td>
<td>4.34 /4.38</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>PA3</td>
<td>4.82 /5.10</td>
<td></td>
</tr>
<tr>
<td>Perceived playfulness</td>
<td>PP1</td>
<td>5.00 /5.41</td>
<td>5.05/ 5.44</td>
</tr>
<tr>
<td></td>
<td>PP2</td>
<td>5.09 /5.62</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>PP3</td>
<td>5.07 /5.31</td>
<td></td>
</tr>
<tr>
<td>Learning motivation</td>
<td>LM1</td>
<td>4.88 /4.69</td>
<td>4.90/ 5.05</td>
</tr>
<tr>
<td></td>
<td>LM3</td>
<td>5.15 /5.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LM4</td>
<td>5.20 /5.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LM5</td>
<td>4.97 /5.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LM6</td>
<td>4.32 /4.59</td>
<td></td>
</tr>
<tr>
<td>Classroom climate</td>
<td>CC1</td>
<td>4.68 /5.79</td>
<td>4.94/ 5.46</td>
</tr>
<tr>
<td></td>
<td>CC2</td>
<td>4.51 /5.45</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>CC3</td>
<td>4.82 /5.71</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>CC5</td>
<td>5.04 /5.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC6</td>
<td>5.01 /5.45</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>CC7</td>
<td>4.97 /5.60</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>CC8</td>
<td>4.99 /5.66</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>CC9</td>
<td>4.96 /5.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC10</td>
<td>4.99 /5.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC11</td>
<td>5.21 /5.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC12</td>
<td>5.16 /5.12</td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td>Incent1</td>
<td>4.73 /5.69</td>
<td>4.64/ 5.54</td>
</tr>
<tr>
<td></td>
<td>Incent2</td>
<td>4.55 /5.48</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Incent3</td>
<td>4.65 /5.47</td>
<td></td>
</tr>
<tr>
<td>Goal conflicts</td>
<td>GI1</td>
<td>4.58 /4.91</td>
<td>4.75/ 4.96</td>
</tr>
<tr>
<td></td>
<td>GI2</td>
<td>4.75 /4.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GI3</td>
<td>4.94 /5.02</td>
<td></td>
</tr>
<tr>
<td>Risk aversion</td>
<td>RA1</td>
<td>4.00 /4.59</td>
<td>4.34/ 4.56</td>
</tr>
<tr>
<td></td>
<td>RA2</td>
<td>4.74 /4.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RA3</td>
<td>4.47 /4.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RA4</td>
<td>4.17 /4.28</td>
<td></td>
</tr>
<tr>
<td>Learning performance</td>
<td>LP1</td>
<td>4.95 /5.26</td>
<td>4.82/ 5.18</td>
</tr>
<tr>
<td></td>
<td>LP2</td>
<td>4.60 /4.97</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LP3</td>
<td>4.93 /5.34</td>
<td></td>
</tr>
<tr>
<td>Confirmation level</td>
<td>Confirm1</td>
<td>4.72 /4.90</td>
<td>4.64/ 4.71</td>
</tr>
<tr>
<td></td>
<td>Confirm2</td>
<td>4.57 /4.53</td>
<td></td>
</tr>
<tr>
<td>Satisfaction level</td>
<td>Sat1</td>
<td>5.11 /5.41</td>
<td>5.05/ 5.04</td>
</tr>
<tr>
<td></td>
<td>Sat2</td>
<td>5.04 /5.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sat3</td>
<td>5.01 /4.52</td>
<td>**</td>
</tr>
<tr>
<td>Continuance intention</td>
<td>Cont1</td>
<td>4.86 /4.90</td>
<td>4.80/4.89</td>
</tr>
<tr>
<td></td>
<td>Cont2</td>
<td>4.77 /5.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cont3</td>
<td>4.79 /4.79</td>
<td></td>
</tr>
</tbody>
</table>

* P < .05; ** P < .01; *** P < .001
Comparative model verification

To provide an equal basis to compare this study with the student study counterpart (Tao et al., 2009), PLS was adopted in this study. A sample of 58 teachers may be an issue since a general guideline for using PLS requires a sample size of over 100 (Gefen, Straub & Boudreau, 2000). Two reasons can justify the use of PLS in this study. First, Barclay, Thompson & Higgins (1995) suggest a different guideline, stating that a sample size exceeding the maximum number of construct indicators is multiplied by ten. The large constructs of this study, learning motivation and risk aversion, have four indicators implying that a sample size of over forty is adequate for applying PLS in this study. According to Gefen et al. (2000), the alternative for SEM analysis is Linear Regression since it has similar capabilities and requires less sample size. Linear Regression analysis did generate a similar outcome as PLS did; hence, it is safe to present the PLS analyses in this study.

However, this study notes the use of PLS as a research limitation for a small sample size of 58 in the Conclusion. This limitation is difficult to address further in this study because the questionnaire was sent to all 122 teachers identified to have adopted BSGs in Taiwan. Follow up calls were also sent to these teachers twice after the first mailing. We learned that many teachers were no longer using BSGs, or had simply asked their teaching assistants to conduct BSG as an assignment or as a competition without personal involvement, due to the high complexity of game management and time investment required to facilitate the process. As a result, the 58 teachers represent the majority of active teachers using BSGs in Taiwan’s higher education institutes.

Table 3 shows nine hypotheses supported by students and only five hypotheses supported by teachers. This obviously implies a certain level of perceptual difference between students and teachers. Among the 17 hypotheses, teachers and students support four and disapprove seven, and the only hypothesis supported by teachers but not by students is H9. The remaining hypotheses, namely, H3, H11, H14, H15, and H17 are supported by students but not by teachers.

<table>
<thead>
<tr>
<th>Theory</th>
<th>Research Hypothesis</th>
<th>Student</th>
<th>Teacher</th>
<th>Differencea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expectation Confirmation Theory</td>
<td>H1</td>
<td>5.43***</td>
<td>6.04***</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>H2</td>
<td>0.91</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>4.44***</td>
<td>0.66</td>
<td>≠</td>
</tr>
<tr>
<td></td>
<td>H4</td>
<td>16.70***</td>
<td>11.54***</td>
<td>+</td>
</tr>
<tr>
<td>Agency Theory</td>
<td>H5</td>
<td>0.63</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H6</td>
<td>0.60</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H7</td>
<td>0.34</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Technology Acceptance Model</td>
<td>H8</td>
<td>0.04</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H9</td>
<td>0.59</td>
<td>2.76**</td>
<td>≠</td>
</tr>
<tr>
<td></td>
<td>H10</td>
<td>4.02***</td>
<td>1.97*</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>H11</td>
<td>6.37***</td>
<td>1.39</td>
<td>≠</td>
</tr>
<tr>
<td>Learning Factors</td>
<td>H12</td>
<td>1.05</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H13</td>
<td>0.57</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Emotional Factors</td>
<td>H14</td>
<td>4.09***</td>
<td>0.41</td>
<td>≠</td>
</tr>
<tr>
<td></td>
<td>H15</td>
<td>3.39***</td>
<td>0.32</td>
<td>≠</td>
</tr>
<tr>
<td></td>
<td>H16</td>
<td>7.28***</td>
<td>4.37***</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>H17</td>
<td>8.14***</td>
<td>0.03</td>
<td>≠</td>
</tr>
</tbody>
</table>

*a + represents both the student’s and the teacher’s support, − represents both the student’s and the teacher’s disapproval, ≠represents that the support of the students and the teachers are inconsistent

Discussion and alignment

Obvious perceptual gaps exist between teachers and students based on the descriptive analysis and model verification analysis. This work discusses and proposes a possible approach to align the perceptions of teachers and students.
Individual construct perspective

Several possibilities may explain the higher scores of teachers in the descriptive analysis. First, teachers and students in general have different appraisal standards, e.g., for the same theoretical perception level, and teachers tend to appraise one scale higher than students do in practice. Second, teachers may simply over-estimate student perceptions. Considering that the 58 teachers voluntarily adopted BSGs in the first place, and students generally hold a more positive attitude towards educational technology (McGill and Hobbs, 2008; Palmer & Holt, 2009) as pointed out in the Background, this study excludes these two explanations. Instead, we believe that the 58 teachers belong to the very small segment of teachers who have adopted BSGs in Taiwan; hence, they are more optimistic than the rest of the teachers who did not, and those students who involuntarily used BSGs in their classes.

Therefore, to align with students on those five factors with significant differences, the teachers can either passively adjust their appraisals of student perceptions on usefulness, playfulness, classroom climate, incentives, and learning performance, or attempt to lift student perceptions. For the latter, either selecting BSGs or enhancing instructional methods or skills can improve student perceptions. In selecting BSGs, teachers should select those that are generally attractive, easy to use, and playful. These games should be properly facilitated to ensure active interactions between teachers and students, as well as among students who are closely tied to learning objects. These games must also be properly conducted to easily enable teachers to motivate, encourage, and reward their students.

For example, the Beer Game conducted in a General Management course is less appropriate than in a Supply Chain Course for expected learning performance and usefulness. The Beer Game is very easy to use due to its simplicity; in addition, its class competition activity is fun to play and can foster an ideal classroom climate by connecting four students in each supply chain and making them compete for lowest costs at both individual and supply chain levels. The teachers can further facilitate the Beer Game in class to enhance student perceptions on learning performance by conducting many game practices with different scenarios and sharing student experiences in class presentations or forum discussions, and to enhance student motivation by granting simple incentives in all learning activities. The use of effective classroom instructional methods and skills can refer to existing educational theories and practices in the literature, such as summarized in Hagborg (1994) and Kyriakides, Creemers, & Antonious (2009). Teachers can also consult Teaching Resource Centers for customized class designs with appropriate learning activities and techniques (Tao, 2008), or form a faculty learning community to provide members with information and support as they move towards utilizing digital technology tools, learn new skills, and share meaningful instructional practices (Nugent et al., 2008).

Path model perspective

Compared with the student model shown in Figure 1, which has partially connected paths from previous factors to the final intention to continue using BSGs, the teacher model presents a disconnected path with the four hypotheses (H1, H4, H10, and H16) with solid bold lines and another hypothesis (H9) with a thin, evenly spread dotted line (Figure 2). The following derives two consequent alignment approaches:

The first approach assumes H9 is supported by students, such that there is a consistent connected path from Perceived attractiveness to Perceived ease of use (H16), to Satisfaction (H9), and to Intention to continue using (H2) between teachers and students. This is a simple path model for teachers to achieve students’ continued usage by selecting BSGs that are attractive or easy to use. Consequently, students may be satisfied, and thus generate a positive response toward using BSGs in future classes. Teachers face the first challenge of switching student perceptions to this causal relationship (H9), and then, the second challenge to identify easy to use BSGs acknowledged by students.

Nevertheless, the first challenge to switch student perceptions of ease of use, which influences satisfaction, is highly challenged as can be gleaned from King & He’s (2007) quantitative TAM model meta-analysis stating that “the effect of ease of use on behavioral intention is primary through usefulness.” The second challenge may also be difficult due to the average scores of Perceived ease of use (4.69 and 4.86) for both teachers and students, which are relatively low among other constructs in Table 2. According to profile analysis, most of the students already used the most popular and sophisticated BOSS (88.1 percent) and Chain Store Master (34.1 percent) games developed by Taiwan developers. The other simpler games, such as the Beer Game and Retailer Expert, are less popular, due to their orientation to specific business functions and applicability.
On the contrary, in the second approach, teachers can take advantage of what students have positively perceived, particularly $H_6$, $H_{11}$, $H_{14}$, $H_{15}$, and $H_{17}$ (Figure 2, unevenly spread dotted lines). As shown, the path model is connected by three paths from perceived attractiveness to the Intention to continuing use, including perceived attractiveness $\rightarrow$ perceived playfulness $\rightarrow$ satisfaction $\rightarrow$ intention to continuing use, perceived attractiveness $\rightarrow$ perceived ease of use $\rightarrow$ perceived playfulness $\rightarrow$ satisfaction $\rightarrow$ intention to continue using, and perceived attractiveness $\rightarrow$ perceived usefulness $\rightarrow$ satisfaction $\rightarrow$ intention to continue using. It is more flexible for teachers to adapt the more appropriate path among the three path options, compared with the single path in the first approach since teachers can simply ignore their beliefs while motivating students to continue using BSGs with their own beliefs.

In practice, teachers must select BSGs that are attractive and then proceed with the selected path to make it work for their respective students. The easiest path may be through playfulness, since the average scores of Perceived playfulness (5.05 and 5.44) are relatively high among other constructs in Table 2. The second path is through usefulness since the average scores of Perceived usefulness (4.99 and 5.47) and Learning performance (4.82 and 5.18) are much higher than those of Perceived ease of use in Table 2 as mentioned above. Students may have perceived current BSGs as reasonably playful and useful to continue taking classes utilizing BSG activities. To enhance the games’ playfulness, usefulness, and other factors, teachers can facilitate the class design as suggested in Section 5.1.

The rationale for preferring the second approach is similar to what Palmer & Holt (2009) suggested, that the lower level needs hygiene factors in addition to the high-level motivating factors presented in the study of Herzberg (1964) to develop teacher perception on the use of educational technology. In this case, Perceived attractiveness, Perceived playfulness, and Perceived ease of use from the sustained hypotheses by students alone are the hygiene factors that serve to motivate teachers.

Other observations

The 122 teachers identified as using BSGs from over 160 higher education institutes in Taiwan is a very small number. Most non-responding teachers had never used the technology, experienced using it once but became inactive, or became slightly involved by designating this responsibility to their own graduate students. The primary reason behind these hindering factors was little or no support for teachers from the administrative units. This problem occurs in educational technologies with the potential of becoming a school-level application, such as e-learning (Tao, 2008), and is generally worse in BSGs for business-oriented courses or in STR for disabled students.
with learning difficulties (Wald, 2005). Students received better support than the staff in the adoption and use of educational technology for an online learning environment (Palmer & Holt, 2009). Even for STR or BSGs, students received generally better support from teachers or their graduate students.

The two clues mentioned may explain two facts. The first fact is that there had been a lack of support to teachers. For example, only 122 teachers were willing to adopt BSGs. A certain percentage of these teachers failed to either start or continue to adopt this study. The second fact is that better support for students and more positive student attitudes towards educational technology (McGill & Hobbs, 2008; Palmer & Holt, 2009) resulted in a more complete path model of students.

To improve the faculty support issue, the teacher resource center is available as shown in Section 5.1. This should also be an easier solution for Taiwan’s higher education institutes to better utilize existing resources to expand their service scope to other educational technologies.

Conclusions and future work

Based on the sample profile comparison in Section 4.1, the comparative descriptive analysis in Section 4.4, and the comparative model verification in Section 4.6, this study concludes that there are significant perceptual gaps between teachers and students on the continuing use of BSGs. The consequences of this misalignment could be that teachers would not adopt BSGs, stop using BSGs after the initial adoption, or encounter challenges and conflicts during the implementation of BSGs in their classes, as discovered in the survey process of this study.

To align the mismatch, we have suggested some strategies. These consist of three resources to narrow the perception gaps of the five significantly different constructs, and two approaches to align the teacher’s path model by linking it to five sustained hypotheses in the student model. These can motivate teachers to adopt BSGs or facilitate their classes better, which may stimulate teachers to continue using BSGs. In investigating the reasons behind non-responding teachers we learned that the teacher support issue is critical to the actual use of BSGs by teachers. We suggest utilizing existing teacher resource centers established for promoting Taiwan’s e-learning environment as a mutually beneficial solution for teachers to seek help from their schools, and for schools to provide comprehensive services to their teachers, including this particular educational technology of BSGs. In addition to remedying alignment strategies, developing BSGs based on the GOMII theoretical framework by Amory (2007) is critical since a high percentage of educational games may be missing pedagogical foundations (Kebritchi & Hirumi, 2008), and thus cause consequent usage issues.

As for the major contribution of this research, TAM, ECT, and emotional factors were sustained in the final research model. This can be a good starting point to understand further the adoption issues related to BSGs in this teacher-student dynamic in addition to related studies regarding benefits and learning performance. A derived practical implication is that even though teachers confirmed and supported a lesser number of hypotheses in the student model, teachers, in reality, can take advantage of the more positive student perceptions to leverage the adoption or continued use of BSGs in class.

Despite our best efforts in conducting this study, some limitations should be mentioned: the teacher model is constrained by its sample size of 58, whose appropriateness for PLS analysis is only partially supported in the literature, even though this is possibly the primary teachers adopting BSGs in Taiwan. This research only sampled the higher education teachers in Taiwan, thus the research outcomes may not be applicable to other regions or countries, despite their referential value as the first study in continued BSG usage in higher education. Among the three unsustained theories and models, the agency theory failed to establish in this study and Tao et al. (2009) can be a future direction for investigating the conflicting message with the potential relationship between intrinsic motivation and extrinsic reward in digital game-based learning (Huang, Huang & Tschopp, 2010).

References


Vaidyanathan, R. & Rochford, L. (1988). An exploratory investigation of computer simulations, students preferences, and


An Analysis of Students’ Academic Performance When Integrating DVD Technology in Geography Teaching and Learning

C. P. (Christo) Van der Westhuizen*, Carisma Nel and Barry W. Richter
Faculty of Education Sciences, North-West University, Potchefstroom Campus, South Africa // christo.vanderwesthuizen@nwu.ac.za // carisma.nel@nwu.ac.za // barry.richter@nwu.ac.za
*Corresponding author

ABSTRACT
This article discusses the effect of the integration of the Digital Versatile Disc (DVD) as an ICT-variant on the academic performance of full-time geography teacher students enrolled for a Bachelor of Education (B. Ed.) degree at a rural university in a developing country. Action research (which includes both quantitative and qualitative studies), done over three years, attempted to determine if the integration of DVD technology can effectively support the teaching and learning of geography teacher students in a learner-centred learning environment, despite a 55% decrease in contact time, as well as a change of normal contact sessions (lectures) to that of a seminar format. It also explains how the DVD was compiled to suit this DVD method, the DVD method itself and how to employ it effectively in conjunction with seminars. Students’ perceptions about this method were continuously considered and changes were made steered by their need and feedback. Results indicate that this learner-centred DVD method was well received and applied by students and that it did not jeopardise their academic performance.

Keywords
ICT in geography teaching, DVD technology, ICT and developing countries, ICT variants, Multimedia learning, Learner-centred learning

Introduction and theoretical framework
A research question often asked when integrating technology into education is “Are there differences between students exposed to traditional face-to-face instruction and students exposed to technology-mediated instruction in an undergraduate course with respect to academic performance, knowledge of, perceived skillfulness in, and attitude toward a specific subject?” (Alkharusi, Kazem, & Al-Musawai). It also has been documented that the use of technology-based instruction in a social constructivist learning environment might have the potential to improve student academic performance, knowledge, skills, and attitudes (Reeves & Reeves, 2008; Tuty & Klein, 2008).

Research, though limited, shows key benefits for geography educators in practice when utilizing Information and communication technologies (ICT) (Bowles, 2006). Lehtinen (as cited in Houtsonen, 2003) examined the significance of the impact of modern ICT for teaching and learning in geography. He did it by means of meta-analyses and concluded that learners in classes where ICT was used as a teaching aid generally learned more than those in other classes, performed better on average in cognitive tests, learned faster, enjoyed the lessons more, and were in general happier in their academic work. Therefore, it seems that hypermedia (ICT) environments produce better results than conventional teaching environments.

The interactive, user-centred, and open structure of new ICTs, particularly the Internet and mass storage devices, are ideal for the creation of constructivist learning environments (Sharpe, 2000). Traditional lecture methods in geography, according to Agnew & Elton (1998) and Van der Schee (2003), must be replaced with interactive lectures, resource-based learning, independent learning, practical and project work, etc. Within this interactive learning environment the learners are able to construct their own knowledge to ultimately perform better.

The role of the university is also changing and is, therefore, altering higher education’s core production and delivery process, which brings new challenges to the fore (Ryan, Scott & Freeman, 2000; Loing, 2005). Lecturers must, therefore, continuously strive to create learning environments and experiences that will enable students to construct their own knowledge rather than embrace the traditional teaching method of knowledge transferral (Van der Schee, 2003). Students will demand flexible, targeted, and accessible learning methods (Ryan, Scott & Freeman, 2000) and these methods should be thought through very carefully (JISC, 2008). Regarding geography teaching and learning, it is the responsibility of the lecturer to create a learning environment that will challenge the student to learn and perform better (Prosser & Trigwell, 1999). It is essential that students be more responsible for their own learning...
(Mills & Cottell, 1998) and that they achieve the set learning outcomes successfully within corporative and resource-based learning environments.

The shift to a learner-centred teaching approach in higher education inevitably means that universities have to reduce the contact time between students and lecturers (Zollo, 1999). According to Agnew (2001), this asks for more effective time management by lecturers, requiring them to spend more time on activities to generate funds for universities, such as research. Together with optimizing contact time between lecturer and student, Zollo (1999) and Van der Schee (2003) explain that greater emphasis is placed on the use of different forms of delivery, making students more self-directed. According to Waters (2002), the belief is that DVD-based instruction delivers on the objectives by employing engaging media, capitalising on what motivates learners and seizing teachable moments.

**Outcome-based education (OBE) and teaching and learning with ICT**

OBE, introduced in South Africa in 1997, placed emphasis on learner-centred education (S.A. Department of Education, 1997; Corru, 2003). This learner-centred teaching approach is based on the social constructivist theory. Constructivist-based learning environments are thus characterized by problem-solving activities, the provision of stimulating learning environments, cooperative learning, promotion of learning through exploration, and reliable assessment methods (Roblyer, Edwards, & Havriluk, 1997). The constructivist approach aims at enabling learners to manage their own learning. OBE strives to enable all students to attain their maximum learning potential in view of the learning outcomes to be achieved at the end of the education process (S.A. Department of Education, 1997). The learner-centred teaching approach of OBE is particularly viable in an ICT learning environment. The OBE approach should offer a supportive and stimulating environment by means of various ICTs, with sufficient learning time and additional learning experiences for learners, wherein learners can work and progress at their own particular pace (S.A. Department of Education, 1997; Grabe & Grabe, 2004).

With ICT, it is possible for learners to learn at different times, different places and without direct supervision by an educator that consequently impacts on the methodology of teaching and learning (Freeman, 1997; Grabe & Grabe, 2004). There are two ways in which ICT can be integrated in the learning process. ICT may be placed in control of the learner’s activities in the learning process in the form of direct teaching packages, known as the technology-centred learning approach, named by Mayer (2001). These ICTs can be used by learners independently of the educator to develop knowledge, understanding or skills. Programmes may include, amongst others, explanations of ideas (with text, sound, diagrams, animations or video), questioning of the learner, feedback on the quality of the response, remediation material if needed, and progression on the next topic (Underwood & Brown, 1997). Putting the ICT in control of the learner’s activities is attractive in many situations in the classroom as it is able to give more immediate feedback than the educator can give and allows the learner to work independently of the educator, who can then devote more time to working on more difficult subject matter. On the other hand, learners can also be put in control of ICTs in order to enhance their learning and, according to Mayer (2001), it is a learner-centred learning approach. These ICTs, which the learners control, include virtual learning environments and are effective for teaching and learning because they provide tools for educators and learners to help manage learning (Kennewell, 2004). The ICT involved assists educators to structure the curriculum over a period of time, to communicate objectives, to provide multimedia resources, and to access the materials. This ICT also helps with setting and scheduling tasks and supporting learners in working together (Becta, 2003). Although both ways are relevant, this study focuses more on the latter. Regarding geography, ICT and this learner-centred learning approach, Bowles (2006) and Storey (2002) state that educators maximise the impact of using ICTs in geography by:

- being clear as to how the use of ICTs will support lesson objectives;
- using ICTs as a tool, not just as an information resource;
- giving learners greater autonomy in their geographical investigations; and
- incorporating the use of portable ICT equipment in teaching and learning.

In their analysis of learning by means of the GLOBUS interactive CD-ROM environmental education programme, Houtsonen & Rehunen (2000) found that more than half the learners regarded this as an interesting and different mode of studying and valued the illustrative animations and sound effects. Timetabling flexibility and the opportunity to study at one’s own pace were also seen as clear benefits of the programme. GLOBUS proved to be well adapted as a study programme for environmental education because it approaches the subject from numerous perspectives and allows ample scope for learners to think matters over for themselves (Houtsonen, 2003). More recently, Golightly (2008) integrated the DVD in the teaching and learning of map work as a learning support...
medium and found that through proper planning, a learning environment that empowered learners to take responsibility for their own learning was created. None of these studies specifically reported on the academic performance of students over a three-year period when integrating these ICTs into geography teaching and learning.

ICTs in developing countries

The challenge to using ICTs in geography teaching and learning occurs in a world experiencing increasing disparities between the rich and poor, both between and within nations (Mansell & When, 1998; Bowles, 2006). In 2002, only 6.4% of South Africans had access to and used the Internet (S.A. Department of Education, 2004). This grew to 9.1% in 2005 in South Africa (Czerniewicz, 2007), which is better than the percentage of users in Africa in 2007, 3.5% (NACI, 2008). Households with access to personal computers (PCs) made up only 13.6% in 2005 (Czerniewicz, 2007). Furthermore, according to SITES 2006, the percentage of schools with computers increased to 38% in 2006, which is still not adequate (Pelgrum & Law, 2008). Telkom in South Africa has failed to broaden the use of landlines to the poor and rural areas. Local calls are still not free and South Africans pay as much as 13 times more for telephone costs than their British counterparts for similar services. As well, calls are over 60% more expensive than in Finland (Economist, 2004). Wireless broadband is currently expensive in limited areas and rural areas will have, for some time to come, Internet access only through dial-up links (Telkom, 2005; Hutheesing, 2005; Gruman, 2005).

The above limitations should not prevent the educator from planning to integrate ICTs, and considering alternative applications of ICTs, for example, the gathering of material on disk from external sources (Bowles, 2006). According to De Moura (1999), there are ICT variants that can be used as alternatives to computers and the Internet, like television broadcasting to improve access to education via ICTs. One of the solutions for sufficient ICT support in teaching and learning for developing countries is to focus on ICT variants that are affordable and that will sustain movement toward fulfilling development objectives (Mansell & When, 1998). Developing countries, according to Lewis (1999), need to consider alternatives to ICT that maximise the impact of ICT and that entail balancing investment in computers with investment in other technologies that might be cheaper and equally effective. The use of ICT variants must, however, be globally competitive and at the same time be cost effective. The ability, versatility, and low cost of an ICT variant such as the digital versatile disc (DVD) suggest that it can serve this purpose (Waters, 2002).

The DVD as ICT variant to support geography teaching and learning

With the introduction of the DVD in 1999, Crawford (1999) highlighted that it is a much more durable audio-visual storage medium than video cassettes. The storing capacity of a DVD is 4.7 billion bytes of raw data (4.7 GB), seven times the capacity of a compact disc. Besides large storage capacity, other benefits according to Morley (2005) are that it is high-quality medium, interactive across platforms and systems, and that information stored on a DVD is randomly accessible. Furthermore, DVDs are interactive in the sense that the user interacts screen by screen with the information and chooses selections to view repeatedly. A DVD can access the Internet and occupies a lot less shelf storage space when compared to video cassettes, for example.

In the Republic of South Africa, a DVD player that can be connected to a television set already costs less than USD50, a DVD-ROM less than USD30, and a writable DVD less than USD1. A portable DVD player with a lift-up 5.8-inch screen costs just under USD164. The DVD has the ability to combine text, audio, photos, animation, and videos and is therefore a learning support medium that has the power to create high multimedia resources, creating digital excitement, and increasing digital literacy. Therefore, a DVD basically offers similar advantages when compared to the computer as an ICT variant, although certain limitations, such as the lack of interactive programmes, do exist.

Method of research

The purpose of this study was to investigate the effect of the integration of the DVD into geography teaching and learning for full-time pre-service teachers at a rural university in a developing country, and at the same time ensure that the students’ academic performance was not jeopardised by this new and unfamiliar DVD method. It was also
necessary to investigate whether the DVD could assist students in a learner-centred learning environment by
determining the compilation and effectiveness of the information included on the DVD as well as the procedural
nature and effectiveness of seminars, as an alternative to the traditional lecture. The DVD was, for practical reasons,
integrated into the second-year economic geography module (GEOH251). The number of formal lectures (four x 50
minute periods each week for five weeks) was reduced to one 90-minute lecture a week, a reduction of 55% of
contact time. These contact sessions were altered to a typical seminar format suitable for the geography lecture room
(according to principles for the geography seminar of Gold, Jenkins, Monk, Riley, Shepherd & Unwin 1991; Tansley
& Bryson, 2000) in order to focus effectively on higher order learning. Typical seminar functions according to Gold
et al. (1991) are proper exercise control, group discussions, reporting and class discussions, facilitated by the
lecturer. Furthermore, the hard-copy study guide of the module was transformed into a DVD guide (DG) but with the
addition of study material, formal lectures, supplementary explanations of concepts, assignment guidance, articles,
videos, animations, photographs and diagrams, all accessible through “click-on” menu insertions in the text (DG).
Each student was issued a portable DVD player that is both battery and electrically powered, that enabled them to
study at any time and place. The purpose of that was that all the knowledge-level information, directions on how to
prepare properly for the group and class discussions, guidelines for assignments, and what to study for the class tests,
were the students’ responsibility. Learner-centred learning was assumed to take place with the assistance/support of
the DVD.

Design and participants
An action research design was implemented, and both qualitative and quantitative data were collected. The
quantitative component, a cross-sectional study as part of a developmental research method (Leedy & Ormrod,
2005), was used to develop, as part of a larger project, a proposed DVD method for the effective integration of the
DVD in geography teaching and learning (see Table 1). An initial DVD method was developed, based on best-
practice principles, to try to answer to the research aim and was implemented with the 2004 student intake.
Observations by the geography lecturer (also the facilitator and first author of this article), as well as feedback
received from the students through the various measuring instruments, were analysed to bring about the necessary
changes to create and implement as the enhanced DVD method for the 2005 students. Further adjustments were
made according to observations, feedback and recommendations from the students and the lecturer to create the
proposed DVD method, which was implemented for the 2006 students. The 2006 students could comment on the
proposed DVD method but no further recommendations were received. The proposed method was used for two more
years with no need for further adjustments.

<table>
<thead>
<tr>
<th>Time</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Obs</td>
<td>Obs</td>
<td>Tx</td>
<td>Group 2</td>
</tr>
</tbody>
</table>

Source: after Leedy & Ormrod (2005)

Notes:
➢ “Obs” refers to an observation with regard to the variables and achievement of learning outcomes made.
➢ “Tx” refers to the presence of a change (“treatment”) in the programme.

For practical reasons the second-year GEOH251 full-time geography module was identified with the central theme as
economic geography. The entire population of the GEOH251 module of 2004 ($n = 42$), 2005 ($n = 31$), and 2006 ($n =
28$) of the B. Ed (teaching degree) used the DVD method with the seminars.

Instrumentation & data collection procedure

For the quantitative component of the study, two questionnaires were used as well as the final module mark of the
students. For the qualitative component, various data-collection methods were used. The main instruments include
the following:
➢ Questionnaires upon completion of the module for the three year groups to determine how the students
perceived, experienced, and valued the DVD method; what the impact on their academic performance was by
comparison; and how to improve the DVD method ($n = 42, 31, \text{ and } 28$ for the three year groups, respectively).

- Personal interviews conducted with some of the students of the three year groups (2004 $n = 4$, 2005 $n = 4$, 2006 $n = 6$) after the processing of questionnaire data and including students who performed below average, average, and above average. This was specifically to receive their input on how to improve the DVD and seminars for the succeeding year group, the impact of the DVD method on their academic performance, and to clarify some of the findings of the questionnaires.

- A comparison between the averages and standard deviations of the academic performances of GEOH251 and the final results of the seven other geography modules over a two-year period of the geography course for the three year groups, respectively.

- An analysis of the averages (ANCOVA) of the final mark of the students of 2003 (traditional lectures only) and 2004–2006 (with the DVD method).

- Other various qualitative data collection methods included recordkeeping (monitoring) of contact and group session activities, informal conversations, and observations of the lecturer.

**Treatment**

**Creation and compilation of the information on the DVD and routing through it**

A technical support team employed by the faculty helped with the production of the DVDs and compiled the information in a certain order, as indicated or required by the lecturer. Figure 1 illustrates the easy manoeuvrability of the DVD player and explains the DVD compilation as well as routing through the information on the DVD study guide (DG). Each block represents an example of a DVD screen and the type of information to expect on it.

Figure 1. Using the information on the DVD and manoeuvrability of the DVD player
The core text, the study guide, now called the DVD guide (DG) in column 1 (fig. 1), contains for example the set module outcomes, learning unit outcomes, study guidelines, exercises, assignments, group activity themes, and possible test and examination questions, similar to the hard-copy study guide. See Figure 2 for an example of a screen shot of this core text. It must be noted that in order to make the text more legible on the DVD player’s screen, it has been spread across a number of screens, which are then fast and easy to read. One can easily move forward and backwards from screen to screen by pressing the “next” or “back” button on the remote control or DVD player. This can mean that one A4 text page, for example, is stretched across four screen pages. Although it could then suggest that the DVD is “text-heavy,” the text only comprises approximately 10–15% of the available information. The larger percentage of information on the DVD is found in the “supportive inserts” as well as in the “additional inserts.”

“Supportive inserts” (column 2, fig. 1) are accessible through numbered menus, for example, (1), (2), etc., integrated in the core text (DG). This includes lectures, videos, and animations to support the module content and lectures, as well as guidelines on tasks, assignments, etc. by the lecturer that normally would happen during class time (see for example screenshot shots in figures 3a and b, which are accessible from the DVD screen in figure 2. Then, from the supportive inserts, relevant “additional inserts,” which contain additional information such as pictures, audio inserts, videos, and films, all to support the supportive inserts can also be accessed through available menus (for e.g., fig. 4 from fig. 3b). These menus in the core text and supportive inserts are on the appropriate DVD screen pages and can be selected, as explained above. Students have the option to choose the available supportive and additional information, or to continue with the core text (compare fig. 1 with screenshot examples in figures 2, 3, and 4).

All text, information, and inserts can be played repeatedly according to the need and learning pace of the student. Using the DVD means that students can now manage their own study time by choosing, for example, the time they wish to listen to the lectures or study a video-clip on a topic not yet mastered.

Translation of screen:
- Which factors eventually lead to the choice of minefields and how different qualities of ore are dealt with.
- Design a short learning experience about this topic.

Read GETIS et al. 1996, p. 347 to 353. While you read through this...

VIDEO INSERT (5)
TEXT BOOK INSERT (3) (Getis, et al. 2004)

The language of tuition is Afrikaans—one of the official languages in South Africa.
Findings and discussions

There were some concerns and expectations regarding the academic performance of the GEOH251 students when integrating the DVD method in geography teaching and learning. It was essential for this study that the results of the GEOH251 were not lower than their results in other subjects of the B. Ed. degree programme or other modules in the geography course. We would be satisfied if the students’ marks stayed, in comparison, more or less the same, given the fact that contact time was 55% less than normal and a different, seminar format was used. If their marks were deemed to be better in comparison, then this research experiment can be seen as successful.

The best way in our view, considering the available data, to determine whether the student’s academic performance had been jeopardised in any way, was to (1) compare the final results of the GEOH251 module with their other subjects’ modules in the B. Ed. degree; (2) compare the GEOH251 module with their other seven geography modules completed over the first two years of the course, and; (3) compare the GEOH251 marks of the 2003, 2004, 2005, and 2006 year groups with one another.

Academic performance compared to other subjects and student perceptions

From the questionnaires, more than 75% (2004 and 2005) and 70% (in 2006) of the students’ marks were equal or better than their marks of other second-year subjects, and as many as 84%, 76% and 80% of all three-year groups, respectively, were surprised with their marks, because they had done better than they had anticipated. As John* commented during the interviews (*pseudonyms used for all names): “. . . it definitely benefits your academic achievement. I think the days of plain classes are of the past and it will become more technology directed and it is a faster way of learning.”

The DVD consists of valuable attributes for students that helped them to achieve better marks for assessments. As George explained during the interviews regarding the advantage of the guidelines by the lecturer as they appear on the DVD:

Yes, it helped a lot because then you have that surety, OK, this is what being expected from you and this is what you are going to look for, and this is how marks will be given, and then you know how you must do it, and then you know you will get good marks for it.

George added that

You can deliver better quality work because you know exactly what to do and we just got higher marks for the assignments, and it saved plenty of time, because then it wasn’t necessary for us to figure out what we have to do.
Regarding the explanations by the lecturer, most of the interviewed students felt more or less the same but this is best put in the words of Charlie: “You can actually physically go back to the work that you have done before and listen to the explanations again just before bigger tests, that helped in getting better marks.”

More than 70% of the students of all three groups reported that the DVD has the potential for more effective use of academic time. All the students (100%) in 2004 and 2005 and 93% in 2006 agreed that the DVD method called for more self-discipline and time management. Furthermore, given that the students are aware of the fact that they live in a developing country, 84% in 2004 and as high as 92% in 2005 and 2006 felt that the DVD is a good alternative to the computer. In 2005, 89% of the students and 93% in 2006 were satisfied with the balance between the independent learning required by the DVD and the seminars.

The DVD-module vs. other geography modules

Students were asked in one of the questionnaires to indicate whether their geography marks were worse, equal to, or better than their marks in other completed geography modules thus far. The response was that 81% of the students in 2004 and more than 75% in 2005 and 2006 indicated that their marks were equal or better.

To mathematically determine whether the academic average of the GEOH251 students of 2004, 2005, and 2006 correlates with the academic averages of all their other completed geography modules of their first two years of the geography course, Table 2 was compiled. It represents the academic averages of the first eight modules of the geography course, the standard deviation of the students’ marks in each module as well as the effect sizes of the academic averages of each module (per year group) compared to the GEOH251 module in which the DVD was integrated. Given the relatively small class sizes, only one class per year group, and timetable issues, it was impractical to use a control group. When no control group exists, the division by $S_{max}$ (see formula below) gives rise to a conservative effect size in the sense that a practically significant result will not be concluded too easily (Ellis & Steyn, 2003). This enabled a comparison between the academic averages of the GEOH251 module that used the DVD method and the other seven modules of the same year group that did not integrate the technology. Figure 2 was compiled from the mean average (%) data of the eight geography modules in Table 2 to compare and to reflect if any abnormalities exist. Table 2 and Figure 2 give an indication of the range in which the GEOH251 averages of 2004, 2005, and 2006 fall. Effect size ($d$) is subsequently calculated to compare the averages and standard deviations of the different modules with the averages of the GEOH251 module using the following formula:

$$d = \frac{|\bar{x}_i - \bar{x}_j|}{S_{max}}$$

where $|\bar{x}_i - \bar{x}_j|$ is the difference between the averages of the two samples ($\bar{x}_i$ and $\bar{x}_j$) without taking the sign into consideration and $S_{max}$ equals the maximum of the two sample standard deviations $S_i$ and $S_j$ (Ellis & Steyn, 2003).

<table>
<thead>
<tr>
<th>GEOH-modules</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>131</td>
<td>61.7</td>
<td>59.5</td>
<td>53.1</td>
</tr>
<tr>
<td>141</td>
<td>62.0</td>
<td>56.2</td>
<td>59.7</td>
</tr>
<tr>
<td>151</td>
<td>58.1</td>
<td>53.9</td>
<td>60.0</td>
</tr>
<tr>
<td>161</td>
<td>56.9</td>
<td>60.0</td>
<td>55.6</td>
</tr>
<tr>
<td>231</td>
<td>61.0</td>
<td>56.1</td>
<td>63.3</td>
</tr>
<tr>
<td>241</td>
<td>56.8</td>
<td>55.6</td>
<td>56.6</td>
</tr>
<tr>
<td>251</td>
<td>65.4</td>
<td>67.5</td>
<td>66.0</td>
</tr>
<tr>
<td>261</td>
<td>64.9</td>
<td>58.4</td>
<td>61.1</td>
</tr>
</tbody>
</table>

* = Medium effect size  
** = Large effect size  
□ = No value for ($d$) is indicated in this case since all modules are compared with GEOH251
From Table 2, the better average of the GEOH251 module in all three groups resulted in the majority of medium \((0.5 \leq d < 0.8)\) and large \((d \geq 0.8)\) effect sizes. Differences of \(d \geq 0.8\) are considered to be of practical significance (Steyn, 2000). These effect sizes give a clear indication that GEOH251 in practice has a greater average and that the students’ academic performances, when integrating the DVD method, were not jeopardized but show a practical significant better average for the 2005 and 2006 groups.

**Figure 2. Averages of geography modules (over two years) for the second-year students of 2004, 2005, and 2006**

According to Figure 2, the assumption can be made that the average of the GEOH251 modules for 2004, 2005, and 2006 are in line with the rest of the students’ geography modules over the first two years of the course. The averages of the eight geography modules are all more or less 60%. It can be noted that the GEOH251 averages for all three year groups have a slightly better average than the rest of the modules across the two years for each year group. Another analysis of the students’ academic performance was, in our view, necessary to further evaluate the success of this method.

### The DVD method and lectures only

**Table 3. Average M-score**, average percentage, and adjusted averages and effect sizes of GEOH251 for 2003–2006

<table>
<thead>
<tr>
<th>Year group (Method of teaching)</th>
<th>Average M-score**</th>
<th>Average in % before adjustment</th>
<th>*<em>ANCOVA</em> Adjusted Averages in %</th>
<th>Effect size ((d)) compared with 2003 (Lecture only compared with the DVD method)</th>
<th>Effect size ((d)) compared with 2004 (DVD method compared with the DVD method)</th>
<th>Effect size ((d)) compared with 2005 (DVD method compared with the DVD method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003 (Lecture only)</td>
<td>12.2</td>
<td>65.6</td>
<td>67.4</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2004</td>
<td>12.3</td>
<td>63.7</td>
<td>65.4</td>
<td>0.23</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2005</td>
<td>13.9</td>
<td>67.4</td>
<td>67.5</td>
<td>0.02</td>
<td>0.25</td>
<td>–</td>
</tr>
<tr>
<td>2006 (Integration with the DVD)</td>
<td>18.1</td>
<td>69.6</td>
<td>66.0</td>
<td>0.16</td>
<td>0.07</td>
<td>0.18</td>
</tr>
</tbody>
</table>

* ANCOVA √MSE = 8.58 for all four year groups
** M-score is determined by allocating a value to the Grade 12 letter grades of each student (maximum six subjects). Each symbol represents a value, that is, an A = 5, B = 4, C = 3, etc., for subjects on the higher grade and for the standard grade an A = 4, B = 3, etc. The higher symbol of the two compulsory languages is multiplied by two. The highest M-score a student can achieve is 35.

To determine and evaluate the academic performance of the GEOH251 students with the integration of the DVD method, an analysis was made of the averages ANCOVA of the final mark of the students of 2003, 2004, 2005, and 2006 in correlation with their average M-scores** (see description below Table 3 for M-score calculations). The
average M-score of the students and the class average of this module of each of the four year groups were compared. The academic average is adjusted according to the M-scores to ensure a fair comparison is made between the academic performance of the different year groups, as shown in Table 3. This enables a comparison between the averages of the three year groups with one another as well as a comparison with the 2003 year group (traditional lecture only) with each of the three year groups, respectively, when the DVD method was in fact implemented.

Effect sizes ($d$) (Cohn’s Criterion) were then used to determine any significant differences between the module marks of the different year groups and are also shown in Table 3. Effect size ($d$) was calculated using the following formula:

$$d = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{MSE}},$$

where $|\bar{x}_i - \bar{x}_j|$ is the difference between the averages of the two samples ($\bar{x}_i$ and $\bar{x}_j$) without taking the sign into consideration and $MSE$ is the mean square error of analysis of covariance (ANCOVA) (Ellis & Steyn, 2003).

The following guidelines were used in the interpretation of the effect sizes: (a) Small effect: $0.2 \leq d < 0.5$; (b) Medium effect: $0.5 \leq d < 0.8$; and (c) Large effect: $d \geq 0.8$. Differences of $d \geq 0.8$ are considered to be of practical significance (Steyn, 2000).

The results in Table 3 further indicate that there are no practical significant differences among the adjusted averages of the different year groups. Although the DVD method was used in 2004, 2005, and 2006, it did not result in poorer academic performances in this module. (It must be noted that, although the module outcomes of the module are the same for all the year groups and all groups had the same lecturer, some variables among the different year groups do exist, for example, different students, different examination papers, etc.). What is, however, noteworthy is the fact that these academic results were obtained with the integration of the DVD method and the change and reduction of contact time that would seem to create more time for research.

The findings of the above three sub-sections also correlate with studies that have found that student-centred approaches may improve student motivation and academic success (e.g., Prendergast, 1994; Lonka & Ahola, 1995). As student Hesti* commented: “The use of the DVD is a great advantage, you can work through the work on your own time on your own pace, go back if you wish, because nothing chases you, so, you then also only work better.”

Interestingly, in conclusion, the majority (over 85% in all three year groups) of the respondents indicated that they are of the opinion that the learning outcomes of GEOH251 were successfully reached and predicted that the DVD teaching method could enhance future achievement.

Conclusions and recommendations

In conclusion, this study indicates that the majority of the students were positive regarding the integration of the DVD in geography teaching and learning. It answers to the learner-centred teaching and learning principles of OBE as well as to the principles on how to integrate ICT effectively in teaching and learning.

One of this method’s primary advantages is that it can be successfully applied in geography classrooms with no access to electricity. Furthermore, the students emphasised that the multimedia and learning content on the DVD (text, lectures, video and audio inserts, etc.) as well as the use of the DVD player, helped foster meaningful learning in economic geography. Moreover, geography students mastered the set outcomes at their own pace without their academic achievement being jeopardised—in fact, to actually perform better.

It is recommended that further studies on the kind of guidance regarding time management and learner support are necessary when using the DVD technology to reduce contact time and promote learner-centred learning. The possibilities of using this DVD method for off-campus/distance learning and to design more collaborative activities to fully utilize the strengths of the DVD can also be further researched. It is further recommended that such a project be continued and expanded to other higher institutions and schools, especially in developing countries struggling with Internet connections, computers, and electricity supply, as well as to other subjects and learning areas.
References


Investigating Learner Affective Performance in Web-based Learning by using Entrepreneurship as a Metaphor

Ming-Chou Liu and Ming-Hsiao Chi

Department of Curriculum Design and Human Potential Development, National Dong-Hwa University, Taiwan // 1Institute of Learning Technology, National Dong-Hwa University, Taiwan // liu@mail.ndhu.edu.tw // 0214michelle@gmail.com

(Submitted June 30, 2010; Revised July 27, 2011; Accepted September 8, 2011)

ABSTRACT

In the era of the Internet, factors which influence effective learning in a Web-based learning environment are well worth exploring. In addition to knowledge acquisition and skills training, affect is also an important factor, since successful learning requires excellent affective performance. Thus this study focuses on learners’ affective performance in Web-based learning. An initial description of the personality traits relating to entrepreneurship and Web-based learning was formulated and then examined with a Delphi survey. The findings show that certain personality traits relating to entrepreneurship are also associated with successful learning in Web-based learning. This study identified four dimensions and ten components which together constitute the metaphor of entrepreneurship in Web-based learning.

Keywords

Web-based learning, Affective performance, Entrepreneurship, Delphi survey

Introduction

As the Internet is already widely used in education, learners have many opportunities to engage in Web-based learning (Engelbrecht, 2005). Over the past decades, researchers have studied such learner characteristics as preferences, perceptions, beliefs, attitudes and self-efficacy in relation to information technology and Web-based learning (Liaw, Chang, Hung, & Huang, 2006; Yang & Tsai, 2008). One of the important challenges of the Web-based learning process is arousing learner motivation (Bento & Schuster, 2003). While most discussions have focused on how to improve online learner’s knowledge and skills (e.g., Sendag & Ferhan, 2009), the importance of affective dimension is gradually receiving more attentions (e.g., Wang & Reeves, 2006; Shen, Wang, & Shen, 2009).

The actual presentation of the affective dimension in learning is the spirit. Barsade, Brief, and Spataro (2003) contemplate affect as a trait which is a stable and long-lasting tendency. If there is a spiritual metaphor for learners to emulate, which indicates the affective contents in the web-based learning environment, there would be a positive effect for the development in online learning, as what Bangert-Drowns & Pyke (2001) mentioned, truly engaged learners are behaviourally, intellectually, and “emotionally” involved in their learning tasks. According to Erkkilä (2000), in addition to its common application to commercial activity, the term of the entrepreneurship can also be applied to personality traits. That is, by taking “entrepreneurship” as a spiritual metaphor, making this term as a learning model is possible in the learning process based on our consideration that a successful learner is one who 1) actively engages in and focus on the learning task, as an entrepreneur actively managing his business; 2) accumulates experience to facilitate his learning, as an entrepreneur accumulating the total capital to grow and expand his business; 3) strives for learning achievement, as an entrepreneur striving for making financial profits; 4) trains himself to be creative in one’s learning process, as an entrepreneur displaying the innovative ideas while running his business; at last, 5) shares his achievements with peers, as an entrepreneur sharing the entrepreneurial achievement with the public and benefiting the society.

This entrepreneurial metaphor might provide a learning model for learners, and some of the traits might equip learners the attitude for achieving successful learning. Take the project-based learning in web-based learning for example, it requires learners investing lots of time and efforts to engage in (Krajcik, Blumenfeld, Marx, & Soloway, 1994), which easily leads learners to their abstention and distraction if lacking a positive attitude and perseverance. With the accentuation upon the entrepreneurial metaphor, an online learner must take the responsibility for success and failure in his or her learning, devoting him- or herself to make it as a successful entrepreneur. After all, Shane, Locke, and Collins (2003) have concluded that technology could be useless without a proper policy. In terms of web-
based learning, a well-designed system or activity might be useless without learners’ determination and exertion in their independent learning.

Finding a suitable metaphoric term is simply a beginning, for identifying the specific and helpful traits are even more important. Compared to the traditional learning context, the change of learning style in web-based environment is mainly based on the different access to the media, the taught knowledge and learning itself are of little difference. In short, developing a future application for cultivating relevant attitudes to the spirit of entrepreneurship and constructing the specific entrepreneurship traits extracted from traditional practices could be possible at the present stage. In the traditional contexts, some individual traits and attributions in learning, such as the motivation, engagement, etc., have been supported by studies that they are highly related to the achievement in a successful learning (e.g., Lim et al., 2006; Conrad & Donaldson, 2004). Therefore, the entrepreneurial spirit generalized and composed in this study might suggest a more efficient and better learning achievement for learners in web-based learning.

Also, the core idea of this study is primarily that online students’ effort as well as their management to their learning should be like entrepreneurs’ engagement in managing their business. By cultivating the entrepreneurship spirit, learners can develop a positive and active approach in their learning instead of seeing learning tasks as unnecessary homework assignments or unwanted burdens. Providing a model for learners to imitate is an encouraging way to improve learner-centered approach within web-based learning (ChanLin, 2008), especially when web-based learning becomes of heavily learner-centered, emphasizing the pervasive and personal learning (Shen et al., 2009). The purpose of this study is to clarify these traits and the related web-based implications for students to emulate in the web-based learning activities.

**Literature review**

**Entrepreneurship**

In the eighteen century, French business leader Richard Cantillon ‘described entrepreneurship as “undertakers” engaged in market exchanges at their own risk for the purpose of making a profit’ (as cited in Roberts & Woods, 2005, p. 46). Schumpeter (1975) proposed a definition of entrepreneurship which focuses on innovative behaviour and consists of five components: developing new and innovative products; proposing new forms of organization; exploring new markets; introducing new production methods; and searching for new sources of supplies and materials. Williams (1981) indicated that entrepreneurship is an innovative process and has a high correlation with economic growth. The definition of ‘entrepreneurship’, however, has developed over time. Timmons (1994) proposed that entrepreneurship centres on creating things of value, a process of making or gaining opportunities, and then developing them. He also proposed that entrepreneurship is a model for thinking, deducing, and acting. Moreover, entrepreneurship is becoming an important part of political and educational programs in many countries around the world (Faltin, 2001).

Carland, Hoy, Boulton, and Carland (1984) identified a large number of entrepreneurial trait characteristics, including: risk tolerance, independence, innovation, sense of responsibility, need for achievement, aspiration, self-confidence, strong interpersonal relationships, communicative competence, strong judgment of risks, aggression, motivation, a need for power, self-management, creativity, zealoussness, ambition, positive reaction to failure, and a willingness to accept challenge.

The trait characteristics identified by Gibb’s (1990) seminal work as being closely related to entrepreneurship include: initiative, strong persuasive powers, moderation, flexibility, creativity, independence, problem-solving ability, a need for achievement, imagination, a strong belief that one controls one’s own destiny, leadership skills, and hard work. More recently, Ward (2005) found the following personality traits to be closely associated with entrepreneurs: risk tolerance, tolerance for uncertainty, vision, capacity to inspire, creativity and innovation, a high internal locus of control, emotional stability, resilience and tenacity, self-awareness, self-confidence, high energy, achievement orientation, a proactive attitude, desire for autonomy, flexibility, initiative, assertiveness, and commitment to others.
Web-based learning and the entrepreneurial spirit in Web-based learning

To cultivate learning competence and apply their knowledge to modern society, students have to construct knowledge and develop skills through the “learning by doing” and “thinking in action” approach, an ideal that is receiving more attention since the advent of Web-based learning. Web-based learning encompasses a variety of approaches and models, each of which requires learners invest considerable amounts of time and effort. For instance, the progressive learning project with a final target, a constructivism-oriented learning method, provides learners with a complicated and authentic project in order to teach such skills as designing, planning, data collection, problem solving, decision making, perseverance, and presentation of results.

Web-based learning is a long-term learning process whose outcome largely depends on the degree to which the learner possesses and applies the trait characteristics of a successful entrepreneur. Taking Gibb’s (1990) ‘hard work’ and Ward’s (2005) ‘achievement orientation’ as examples, Web-based learners have to dedicate themselves to their project and expect to succeed, just like an entrepreneur does when initiating a business venture.

As mentioned above, researchers have identified many specific trait characteristics of entrepreneurs, and according to Gibb’s (1990) definition, someone who presents a majority of these trait characteristics can be called an entrepreneur. This study proposes a framework consisting of the four main dimensions used to describe the metaphorical entrepreneurship of learners in Web-based learning (shown as Figure 1). The classification and the sequence of the four dimensions in this framework are based on the Stage Model of Information Processing from information procession theory (Huitt, 2003), and such application of input-process-output structural framework in ICT teaching has been generally accepted. Above all, a core idea of ‘life cycle’ is provided to this framework, encouraging one to develop his/her spirit of entrepreneurship through a constant learning cycle of practice, application and modification. Every components included in the four dimensions are mainly derived from the descriptions of entrepreneurship connotation in literatures written by Carland et al. (1984), Gibb (1990), Foster & Lin (2003), and Ward (2005), etc., and they are identified and confirmed through expert panel of Delphi technique in this study.

**Figure 1. A framework of entrepreneurship in Web-based learning**

The first dimension is ‘vision’. In entrepreneurship this refers to discovering a promising business opportunity and starting an enterprise in order to gain a profit (Gibb, 1990; Ward, 2005); in Web-based learning this refers to discovering an excellent learning opportunity and taking advantage of it in order to increase one’s knowledge. A learner with vision knows what needs to be learned and why it is important. The most important component of vision is initiative, and learners with initiative are expected to have a considerable pioneering and unique vision seeking a learning opportunity and learning from it. With an active attitude, learners must have ambition for success, and possess a high need for achievement (Hornaday & Aboud, 1971; Gibb, 1990). That is to say, after choosing a certain
goal, learners need to have a strong desire for achievement; this is what Ward (2005) called ‘achievement orientation’.

The second dimension is ‘action’. In entrepreneurship this refers to work diligently, moderately and effectively, while in Web-based learning it means to learn diligently, moderately and effectively. After clear understanding of the learning task, learners not only have to make appropriate and focused strategies, but also should be able to complete their learning tasks in each stage of web-based learning activity by following the prescribed order. They need to spend time in practice and exercise moderation in their learning endeavours. In addition, “action” and the following dimension “thinking in action” described below are intertwined and mingled with each other.

In the third dimension “Thinking in Action,” one’s working and short-term memory which promotes one’s self-examination and work review is identified in ‘thinking.’ In other words, this ‘thinking’ leads one to be of a dynamic state not only in action but also in thinking. There are three components in this dimension. The first is creativity. According to a report on creativity and cultural education by the National Advisory Committee, creativity includes imagination, goal orientation, and originality. It also contains the meaning of fluency, flexibility, originality, and elaboration as well (Guilford, 1977; Torrance, 1974; Gibb, 1990). All of which are necessary if learners in an E-learning environment are to produce creative and valuable results. In ordinary learning process, to maintain the fluency and flexibility of concept development and induction, one needs abundant ideas and imagination to solve pending problems in work. Learners should find creative ways to solve difficulties and make their learning process much more fluent. The second component is leadership. It means possessing strong communication skills. Learners with leadership interact well with peers, encourage others to engage in cooperative learning, and know how to persuade them to accept good ideas about a new idea. All of which facilitate the kind of peer interactions which result in highly effective teamwork (Gibb, 1990) and successful learning. This also entails possessing a strong belief that one controls one’s own destiny (Gibb, 1990), independence, and autonomy (Hornaday & Aboud, 1971; Gibb, 1990; Ward, 2005). The third component is autonomy, which refers to one’s good self-management (Timmons, 1978; Welsh & White, 1981). In entrepreneurship, this means one can work in a self-regulated way. Time management, using immediate feedback to modify working methods, and immediate solutions on problems are all included. In web-based learning, self-managed learning is a process in which the learner is responsible for identifying what is to be learned. The learner is responsible for evaluating the effectiveness of the learning activity and whether it is relevant to the objective (Guglielmino & Guglielmino, 2001).

The fourth dimension is ‘feedback’. In entrepreneurship this refers to evaluating self-achievements in a positive way, promoting public welfare, donating money, giving speeches, and other ways of benefitting society; in Web-based learning, this refers to evaluating achievements, presenting learning results, sharing one’s experience, and giving feedback to peers. These contributions of ‘feedback’ are especially highlighted in effective virtual learning community, for they are not only of indispensable for students in the age of the internet (Palloff & Pratt, 1999, 2003) but also of potential for cultivating one who is likely to make contribution to the society. After completing the learning tasks and achieving the learning goals, learners should confidently take stock of their achievements and experiences, and also share these with others.

Methodology

The Delphi technique uses a series of questionnaires to collect data from a panel of selected subjects and is well suited for consensus-building (Dalkey, 1969; Linstone & Turoff, 1975; Martino, 1983; Young & Jamieson, 2001). In this study we used the Delphi technique to collect opinions from experts to form a consensus on using entrepreneurship as a metaphor for Web-based learning.

The first step consisted of conducting a literature review in which related data was collected and analyzed in order to determine the trait characteristics of entrepreneurship that have a bearing on Web-based learning. In the second step, a panel was formed to discuss and determine the preliminary elements, and the experts were asked to express their degree of agreement on the four dimensions of entrepreneurship in Web-based learning, as shown in Table 1. In the third step, the qualitative and quantitative methods were used to integrate the opinions of twelve experts to form a consensus on the trait characteristics of entrepreneurship that have a bearing on Web-based learning.
Table 1. Preliminary elements of entrepreneurship in Web-based learning

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Entrepreneurship in Web-based learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td></td>
</tr>
<tr>
<td>1. Initiative</td>
<td>1-1 Comprehension of the learning goal</td>
</tr>
<tr>
<td></td>
<td>1-2 Clear sense of purpose</td>
</tr>
<tr>
<td></td>
<td>1-3 Proactive thought and action</td>
</tr>
<tr>
<td></td>
<td>1-4 Willingness to undertake new endeavours</td>
</tr>
<tr>
<td>2. Need for achievement</td>
<td>2-1 Strong motivation and ambition</td>
</tr>
<tr>
<td></td>
<td>2-2 Self-confidence</td>
</tr>
<tr>
<td></td>
<td>2-3 High expectations for learning</td>
</tr>
<tr>
<td>3. Diligence</td>
<td>3-1 Endeavouring to finish each learning task</td>
</tr>
<tr>
<td></td>
<td>3-2 Investing time in learning tasks</td>
</tr>
<tr>
<td></td>
<td>3-3 Ability to achieve concrete results</td>
</tr>
<tr>
<td>4. Moderation</td>
<td>4-1 Making appropriate decisions about learning tasks after deliberating</td>
</tr>
<tr>
<td></td>
<td>4-2 Completing learning tasks step by step</td>
</tr>
<tr>
<td>5. Effectiveness</td>
<td>5-1 Problem solving ability</td>
</tr>
<tr>
<td></td>
<td>5-2 Learning effectively</td>
</tr>
<tr>
<td></td>
<td>5-3 Adopting appropriate learning strategies</td>
</tr>
<tr>
<td>6. Autonomy</td>
<td>6-1 Finding problems and asking questions</td>
</tr>
<tr>
<td>7. Creativity</td>
<td>6-2 Time management</td>
</tr>
<tr>
<td></td>
<td>6-3 Learning automatically</td>
</tr>
<tr>
<td></td>
<td>6-4 Learning accountability</td>
</tr>
<tr>
<td>8. Leadership</td>
<td>7-1 Fluency</td>
</tr>
<tr>
<td></td>
<td>7-2 Originality</td>
</tr>
<tr>
<td></td>
<td>7-3 Flexibility</td>
</tr>
<tr>
<td></td>
<td>7-4 Expressive ability</td>
</tr>
<tr>
<td>9. Self-confidence</td>
<td>8-1 Strong interpersonal skills with peers</td>
</tr>
<tr>
<td></td>
<td>8-2 Strong leadership skills in cooperative learning</td>
</tr>
<tr>
<td></td>
<td>8-3 Ability to persuade peers to accept good ideas</td>
</tr>
<tr>
<td>10. Sharing of achievements</td>
<td>8-4 Expressive ability</td>
</tr>
<tr>
<td></td>
<td>9-1 High self-evaluation of learning results</td>
</tr>
<tr>
<td></td>
<td>9-2 High self-evaluation of working results</td>
</tr>
<tr>
<td></td>
<td>9-3 Confidence to enhance outcomes through introspection</td>
</tr>
<tr>
<td></td>
<td>10-1 Ability to share learning experience with peers</td>
</tr>
<tr>
<td></td>
<td>10-2 Ability to present learning results through the internet</td>
</tr>
<tr>
<td></td>
<td>10-3 Ability to enhance outcomes through introspection</td>
</tr>
</tbody>
</table>

The twelve experts of Delphi panel have experiences of promoting web-based learning education and research more than a decade, including eight professors who majored in learning technology and lecturing at universities and four teachers with extensive Web-based teaching experience serving or had been serving in-field in primary schools and junior high schools as information education leader and manager of web-based learning design and promotion.

In each iteration of the Delphi technique, the panel of experts responded to a questionnaire that included the quantitative values of mean, standard deviation, medium, mode and number of respondents. They were also presented with qualitative descriptions of their persistent and unique opinions in the previous round and asked to revise their judgments or to specify the reasons for remaining outside the consensus.

We used the method proposed by Scheibe, Skutsch, and Schofer (1975) to judge whether or not the opinions of the experts had reached stability. Specifically, a change of 15% was taken to represent a state of equilibrium; hence, any two questionnaires that had marginal changes of less than 15% were considered to have reached stability, resulting in the conclusion of the Delphi technique. We executed a total of three iterations of the Delphi technique to reach the final consensus. According to previous studies (e.g. Brooks, 1979; Custer, Scarcella, & Stewart, 1999; Ludwig, 1997), three iterations are often sufficient to collect the necessary information and reach a consensus. We have spent two months on each run of iteration, totally six months to reach a general agreement. Table 2 shows the results of the three rounds of the Delphi questionnaire. In Table 2 “—” indicates deletion of the description, for example, ↓1-3 and ↓4. “*” indicates the addition of an affective description which did not exist in the first round. However, the addition of the description did not guarantee that it would appear in the final affective description.
### Table 2. Statistical results of three rounds of the Delphi questionnaire

<table>
<thead>
<tr>
<th>Criteria</th>
<th>First round</th>
<th>Second round</th>
<th>Third round</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Quartile deviation</td>
</tr>
<tr>
<td><strong>Vision</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-1</td>
<td>4.2</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>1-2</td>
<td>3.8</td>
<td>1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>1-3</td>
<td>3.5</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>1-4</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2-1</td>
<td>4.3</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>2-2</td>
<td>3.9</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>2-3</td>
<td>4.2</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-1</td>
<td>4.4</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>3-2</td>
<td>3.5</td>
<td>1.3</td>
<td>1.5</td>
</tr>
<tr>
<td>3-3</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4-1</td>
<td>3.3</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>4-2</td>
<td>4.1</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>5-1</td>
<td>4.0</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>5-2</td>
<td>3.9</td>
<td>0.8</td>
<td>0.0</td>
</tr>
<tr>
<td>5-3</td>
<td>3.8</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Thinking in action</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-1</td>
<td>4.3</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>6-2</td>
<td>4.3</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>6-3</td>
<td>4.3</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>6-4</td>
<td>4.1</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>7-1</td>
<td>3.8</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>7-2</td>
<td>3.7</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>7-3</td>
<td>3.3</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>7-4</td>
<td>3.3</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td>8-1</td>
<td>4.0</td>
<td>1.1</td>
<td>0.6</td>
</tr>
<tr>
<td>8-2</td>
<td>3.7</td>
<td>1.1</td>
<td>0.8</td>
</tr>
<tr>
<td>8-3</td>
<td>3.6</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-1</td>
<td>4.2</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>9-2</td>
<td>4.1</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>9-3</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>10-1</td>
<td>4.1</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>10-2</td>
<td>3.8</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>10-3</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

### Results and discussion

The four dimensions containing the ten components of affective competence were divided into twenty-eight concrete elements of entrepreneurship in Web-based learning. While the number of components remained the same before and after the Delphi survey, some modifications were made to the number of concrete elements. Specifically, the number of elements in the vision dimension decreased, and the number of elements in the action dimension increased. The number of elements in the “thinking in action” and “feedback” dimensions remained the same. Table 3 maps the relationship between entrepreneurship in business and the metaphorical entrepreneurship of Web-based learning.

### Table 3. Entrepreneurship in business and metaphorical entrepreneurship in Web-based learning

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Entrepreneurship in business</th>
<th>Metaphorical entrepreneurship in Web-based learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vision</strong></td>
<td>1.Initiative</td>
<td>Possessing clear comprehension to the learning goal</td>
</tr>
<tr>
<td></td>
<td>Showing strong enthusiasm in work</td>
<td>Showing strong enthusiasm in work</td>
</tr>
</tbody>
</table>

207
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Entrepreneurship in business</th>
<th>Metaphorical entrepreneurship in Web-based learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Need for achievement</td>
<td>Strong motivation and ambition to the success</td>
<td>Strong motivation and ambition to complete the learning task successfully</td>
</tr>
<tr>
<td></td>
<td>High expectations for profit</td>
<td>High expectations for learning outcomes</td>
</tr>
<tr>
<td>3. Diligence</td>
<td>Endeavouring to accomplish business tasks</td>
<td>Endeavouring to accomplish learning tasks</td>
</tr>
<tr>
<td></td>
<td>Investing time in business tasks</td>
<td>Investing time in learning tasks</td>
</tr>
<tr>
<td>4. Moderation</td>
<td>Making appropriate decisions about business tasks after deliberating</td>
<td>Making appropriate decisions about learning tasks after deliberating</td>
</tr>
<tr>
<td></td>
<td>Able to complete business tasks by following the prescribed order</td>
<td>Able to complete learning tasks in each stage of web-based learning activity by following the prescribed order</td>
</tr>
<tr>
<td>5. Effectiveness</td>
<td>Able to proffering practical thoughts and methods to complete the business tasks with a clear understanding of the task</td>
<td>Able to proffering practical thoughts and methods to complete learning tasks with a clear understanding to the task</td>
</tr>
<tr>
<td></td>
<td>Working efficiently</td>
<td>Learning efficiently</td>
</tr>
<tr>
<td></td>
<td>Taking focused working strategies</td>
<td>Taking focused learning strategies</td>
</tr>
<tr>
<td>6. Autonomy</td>
<td>Finding problems and resolving them right away</td>
<td>Finding problems and asking questions right away</td>
</tr>
<tr>
<td></td>
<td>Using time management to control work time</td>
<td>Using time management to control learning time</td>
</tr>
<tr>
<td></td>
<td>Using feedback as an inspection for the performance review of the business</td>
<td>Using feedback as an inspection for the performance review of Web-based learning process</td>
</tr>
<tr>
<td></td>
<td>Modifying working methods based on business results</td>
<td>Revising learning methods and strategies based on learning results</td>
</tr>
<tr>
<td>7. Creativity</td>
<td>Fluency of concept development, reflection and induction in business management</td>
<td>Fluency of concept development, reflection and induction in the learning process</td>
</tr>
<tr>
<td></td>
<td>Originality in business management</td>
<td>Originality in the learning outcomes</td>
</tr>
<tr>
<td></td>
<td>Diversification toward working ideas and methods and flexibility to solve businesses problems</td>
<td>Diversification toward learning ideas and methods and flexibility to solve learning difficulties</td>
</tr>
<tr>
<td></td>
<td>Elaborate and careful thinking in business management</td>
<td>Elaborate and careful thinking in the learning process</td>
</tr>
<tr>
<td>8. Leadership</td>
<td>Maintain good relationships with co-workers and business associates with strong interpersonal skills</td>
<td>Encouraging others to think in the process of peers interaction</td>
</tr>
<tr>
<td></td>
<td>Knowing how to lead or to direct others to work together while proceeding business cooperation</td>
<td>Knowing how to lead or direct others to engage in learning and profound discussion while proceeding peer learning</td>
</tr>
<tr>
<td></td>
<td>Ability to persuade co-workers or business associates to accept good ideas on working task</td>
<td>Ability to persuade peers to accept good ideas about learning task</td>
</tr>
<tr>
<td></td>
<td>Positive self-evaluation of business results</td>
<td>Positive self-evaluation of learning results</td>
</tr>
</tbody>
</table>
Table 3 shows that entrepreneurship is an appropriate metaphor for Web-based learning. As indicated by the first dimension, vision, effective learners comprehend the learning goal, actively engage in learning, and expect positive results, which can be regarded as having a sense of self-efficacy in education. Self-efficacy is a judgment a person has about himself concerning his ability to deal with an intellectual, social, affective or physical situation (Bandura, 1977). Self-efficacy theory posits a strong connection between beliefs and engagement, so that people with a strong sense of self-efficacy are likely to persist and succeed (Bandura, 1982). Many researchers have suggested that a student’s sense of self-efficacy is a good predictor of academic achievement and motivation (Graham & Weiner, 1996; Pajares, 2003; Pintrich & DeGroot, 1990; Pintrich & Schunk, 1995). The strong relationship between expectations and academic achievement has been well established both theoretically and empirically (Johnson, Livingston, Schwartz, & Slate, 2000; Marzano, 2003).

As indicated by the second dimension, action, successful learners dedicate themselves to their learning tasks. This process is highlighted by the fundamental idea underlying engagement theory: students must be meaningfully engaged in learning activities through interaction with others and meaningful tasks. In principle, such engagement could occur without the use of technology (Kearsley & Shneiderman, 1999). Yet, the essential characteristics of the Internet include communication, immediate feedback, and automatic assessment, all of which facilitate learner engagement and performance.

The third dimension, thinking in action, stresses the development of the learner’s thinking skills, including autonomy, creativity and leadership. The denotation of self-regulated learning exhibits learners’ autonomy. Self-regulated learners continually engage in four activities: planning, organizing, monitoring, and evaluating the learning process (Corno, 1989; Zimmerman & Paulsen, 1995). Being self-regulated entails using all the skills that enable one to meta-cognitively, motivationally, and behaviourally participate in one’s own learning process (Zimmerman, 1986). Self-regulated learners tend to exhibit goal directedness, manage their learning time, meaningfully engage in learning practice, use cognitive and meta-cognitive strategies appropriately, and have a sense self-efficacy regarding their learning tasks. Since thinking directs a learner’s behaviour, learners must hone their thinking skills in order to better direct their learning efforts and attain their learning goals. On the other hand, creativity brings valuable benefits to learners and classrooms even in different educational subjects (e.g., Kaufman & Sternberg, 2007; Piggott, 2007), and the ability to persuade and lead peers also facilitates learners to have successful learning.

As indicated by the fourth dimension, feedback, the purpose of education is pursuing academic achievement and eventually serving society. On the whole, in the process of Web-based learning, learners cultivate the spirit of entrepreneurship. Accordingly, Web-based learning activities should be designed so as to cultivate the personality traits of an entrepreneur in the learner, and also facilitate self-accountability.

Using the traits of entrepreneurship as a metaphor for learning, the two most significant contributions are the following: (1) to provide a concrete prototype for learners, which is related to the existing issues of learning technology. Although this idea is certainly not new, it is a natural conclusion to this study; (2) to utilize the entrepreneurship spirit, in this era of knowledge economy, as a model for the learners’ attitudes toward learning. To quote a well-known aphorism, “Attitude determines altitude.” After all, EQ (Emotional Intelligence) and SQ (Social Intelligence) which have been more often ignored by educator should have played vital and necessary roles as IQ (Intelligence Quotient) in the learning process.
We extract the essence of the traditional learning traits that can be transformed into the metaphorical entrepreneurship traits in web-based environment. In the current situation, however, we don’t expect a website being provided with such a quality of cultivating learner’s spirit of entrepreneurship, entirely or partially, for the nature of the learning traits cultivation is usually non-formal and/or hidden. With the consideration of ‘how to achieve most of components we found through tools/pedagogy of Web-based learning’, we expect this study calling the attention to the practice of entrepreneurship cultivation. Moreover, with a long term perspective, developing rules and rubrics for web-based learning environment in the near future would be capable of being expected as well. As applying the entrepreneurship traits into online project-based learning practically, for example, by connecting the statements of the two dimensions, vision and feedback to encourage learners to conceive and promise the efficacy of learning outcomes and its specific influence upon other learners, or postpone the project schedule and increase the challenge of learning in order to encourage learners to develop a positive attitude toward Web-based learning.

This study starts with the interpretation by the metaphorical entrepreneurship, and investigates learner’s affective attitude in learning. As table 3, we make use of the idea of entrepreneurship borrowed from business management and bring it to the issue of “Metaphorical entrepreneurship in Web based learning.” Indeed, the components in online context generalized by this research are included in those of traditional context; however, it is reasonable for the boundary of the real world and the virtual world has become blurred. Given that the Internet application is usually personal and private from user’s point of view, we suggest learners should face various online learning more with consistent affective attitude which is often neglected and marginalised. “It is time to redress the imbalance by developing theories and technologies in which affect and cognition are appropriately integrated with one another” (Picard, Strohecker, Papert, Bender, Blumberg, & Breazeal, 2004). Meanwhile, through emphasizing the importance of affective domain to web-based learning by presenting experts’ definition and perspectives through Delphi technique, the process of this research could be more significant than the result.

Conclusion and suggestions

With the progress and extensive application of Internet technology, Web-based learning has rapidly developed into an effective educational method. Some scholars have expressed concern that Internet education fails to nurture essential positive personality traits (Bayram, Deniz, & Erdoğan, 2008). This study identifies a number of positive trait characteristics held in common by entrepreneurship and Web-based learning, indicating the cultivation of learner affect in a Web-based learning environment. Entrepreneurship could be a likely metaphor for Web-based learning. Online learners are supposed to be as successful entrepreneurs placing emphasis on the “activeness” of “management” in the four dimensions of learning process. Only online learners with such attitude can be called a successful learner. As entrepreneurs, “who looks upon a business as though it is his own,” (Chang, 2009) they possess the spirit of entrepreneurship and consider the web-based learning tasks as their own enterprise.

Although the components proposed by this study have no specific or direct evidence to support that all the attributes do relate to learning performance, any identified traits in the four dimensions are almost recognized by the teaching and learning field and supported by the related educational literature, and the direct contribution of this study at the current stage is to arouse the attention to embed these components in web-based learning, emphasizing that the spirit of entrepreneurship should be cultivated among web-based learners. Although these traits could also be found in the traditional context for achieving high quality learning outcomes, we expect that the basic metaphor application (suggested as the right column in Table 3) expanded and infused into the network practices in the future, carrying out its actual significance for the web-based learners in their effective performance.

To sum up, Web-based education must include more than knowledge acquisition and skill training; it must also cultivate the learner’s personality and overall maturity. Though web-based and e-learning has been widely used as a common way for learning; however, we regard that the learner’s sense of responsibility for learning is even more essential than any other learning style, and must beyond the technology of the internet and the digitalized information. The entrepreneurship mentioned in this study seems to be more important for its inclusion of the traits highly related to online learning, and this spirit is highly corresponded with the learner-centered idea. Beside the rare investigation on learners’ effective attitudes in web-based learning, we expect that there would be more research and teaching methods make effort to the possible affective attitudes for online learning rather than the theoretical pursuit for knowledge.
For future study suggestions, there are several directions: 1) after we use the entrepreneurial metaphor that could contribute to a successful online learning, the further researches that are likely to prove the concept of the spirit of entrepreneurship could further penetrate into the efficacy and practicability that these traits indeed are effective to learner’s achievement in web-based learning. 2) According to the foregoing specific traits, the expanded web-based activities integrating with the entrepreneurial spirit could be designed to launch a real class. We expect, in addition to the highlight on knowledge acquisition and the skill learning, that there are more designs of web-based activities which help to cultivate the affective traits more.

Acknowledgements

This research was supported by the National Science Council of the Republic of China under contract 96-2520-S-026-001 and 97-2511-S-259-007-MY3, and National Central University’s Learning Technology Project (9-H-FA07-1-4). The author would like to give special thanks to Dr. T. W. Chan.

References

findings and perspectives in entrepreneurship. Brookfield, WI: Avebury.


A User-Centric Adaptive Learning System for E-Learning 2.0

Shiu-Li Huang* and Jung-Hung Shiu1
Department of Business Administration, National Taipei University, New Taipei City, Taiwan // 1Department of Information Management, Ming Chuan University, Taoyuan County, Taiwan // shiulihuang@gmail.com // vita73917@gmail.com

*Corresponding author

(Submitted August 3, 2010; Revised April 20, 2011; Accepted June 28, 2011)

ABSTRACT

The success of Web 2.0 inspires e-learning to evolve into e-learning 2.0, which exploits collective intelligence to achieve user-centric learning. However, searching for suitable learning paths and content for achieving a learning goal is time consuming and troublesome on e-learning 2.0 platforms. Therefore, introducing formal learning in these platforms to provide learning guidance is important. Adaptive learning mechanisms are useful to provide learning guidance based on individual differences. However, most adaptive learning systems provide learning paths and content based on the views of a few designers or experts. To tackle these problems this research proposes a user-centric adaptive learning system (UALS) that uses sequential pattern mining to construct adaptive learning paths based on users’ collective intelligence and employs Item Response Theory (IRT) with collaborative voting approach to estimate learners’ abilities for recommending adaptive materials. The experimental results show that the effectiveness of user-centric adaptive learning is comparable to expert-designed learning and learners are more satisfied and learn efficiently. The guidelines to design e-learning 2.0 platforms are also proposed.

Keywords
Adaptive learning, Concept sequences, E-learning 2.0, Item Response Theory, User-centric learning

Introduction

Accompanying the rapid growth of Web 2.0, e-learning is evolving toward a new trend: e-learning 2.0 (Downes, 2005). Learners share their knowledge, search for the knowledge they need, and decide learning content by themselves through social software platforms. Therefore, e-learning 2.0 is a kind of collaborative and user-centric learning, which is based on collective intelligence rather than a few experts’ knowledge. Some studies propose e-learning 2.0 systems that involve learners in designing, problem solving, or decision making through collaboration and communication tools (Helic, Maurer, & Scerbakov, 2004; Helic, Krottmair, Maurer, & Scerbakov, 2005; Chow, Fan, Chan, & Wong, 2009). In an e-learning 2.0 environment, however, searching and navigating specific knowledge is typically a tedious and time-consuming task (Safran, Helic, & Gütl, 2007). Such a variety and quantity of knowledge content may cause information or cognitive overload. Consequently, learners easily lose their focus and control over their learning processes (Karrer, 2006b).

Adaptive e-learning is able to provide efficient and formal learning by supporting different learning paths and materials to fit learners’ diverse needs and backgrounds (Bra, Brusilovsky, & Houben, 1999; Mallak, 2001; Blochl, Rumetshofer, & Wob, 2003). However, learning paths and content provided by most adaptive learning systems are designed by a few experts and disobey the user-centric principle in e-learning 2.0. We expect that combining user-centric learning with adaptive learning based on collective intelligence will be able to supply learners with more self-control and promote efficient learning in the e-learning 2.0 environment.

This study aims to develop a user-centric adaptive learning system called UALS. This system can be built on e-learning 2.0 platforms such as educational blogs, wikis, or blikis (Huang & Yang, 2009). These platforms enable users to share, create, and collaboratively edit knowledge content. The UALS exploits users’ collective intelligence to dynamically provide adaptive learning paths and materials. This research adopts a collaborative voting approach enabling users to mutually decide material difficulties, and applies sequential pattern mining to extract concept-sequence patterns from user-created resources. The system uses sequential rules to personalize user-centric learning paths and employs Item Response Theory (IRT) to evaluate learners’ abilities and recommends the most suitable learning content to them. To evaluate the UALS, this research will certify that learning guided by users’ collective intelligence is comparable to expert-designed learning.
E-learning 2.0

The main characteristic of e-learning 2.0 is that students can actively control their learning content and direction, which restores learning commands to learners (Downes, 2005; Karrer, 2006a). Corresponding to Web 2.0, Internet users create and distribute content to others by social software, e.g., blogs and wikis. In e-learning 2.0, learning content is no longer produced by instructors or courseware authors and becomes more user-centric, where content is used and created by learners themselves (Downes, 2005). In other words, e-learning 2.0 links learners with other learners, as well as learners and learning resources.

Karrer (2006a) compared the e-learning generations, such as e-learning 1.0, e-learning 1.3, and e-learning 2.0. As e-learning develops, it becomes more user-centric, bottom-up fashioned, and more on-demand. In e-learning 2.0, online learning becomes a platform in which content is created, shared, remixed, repurposed, and passed along rather than a medium in which content is only delivered and consumed. E-learning software becomes a content-authoring tool and not a content-consumption tool (Chow et al., 2009).

Today’s e-learning 2.0 platforms are able to support the creation and sharing of knowledge content and building collective intelligence. Learners can search for knowledge content and decide which content is suitable for their learning. However, searching and organizing suitable content can easily make learners lose their focus on learning. Therefore, how to provide learning guidance based on collective intelligence is an important issue.

Item Response Theory

IRT can measure a learner’s ability on the basis of strict assumptions and mathematical principles (Lord, 1980). The following formula is the one-parameter logistic model (Hambleton, 1985; Horward, 1990):

\[
P_j(\theta) = \frac{e^{(\theta - b_j)}}{1 + e^{(\theta - b_j)}}, j = 1, 2, \ldots, n. \tag{1}
\]

Assume that a learner responds to the \( j \)th item, \( e \) is the mathematical constant 2.71828, \( n \) is the total number of items, \( b_j \) represents the difficulty level of item \( j \), and \( P_j(\theta) \) denotes the probability that the learner can correctly respond to the \( j \)th item with his/her ability \( \theta \). In addition to the parameter item difficulty, the model can further consider the parameters discrimination level of the item and the probability that the learner responds with the correct answer by guess. For simplicity, this research focuses on the one-parameter model. Maximum likelihood estimation (MLE) is usually applied to estimate a learner’s ability based on the following likelihood function (Hambleton, Swaminathan, & Rogers, 1991):

\[
L(u_1, u_2, u_3, \ldots, u_n | \theta) = \prod_{j=1}^{n} P_j(\theta)^{u_j} Q_j(\theta)^{1-u_j}, \tag{2}
\]

where \( (u_1, u_2, \ldots, u_n | \theta) \) is a response pattern that a learner with ability \( \theta \) responds to a set of \( n \) items. An element \( u_j \) (1 \( \leq \) \( j \) \( \leq \) \( n \)) is either 1 or 0 for the \( j \)th item; 1 represents that the learner correctly responds to the item, and 0 represents that the learner incorrectly responds to the item. \( Q_j(\theta) \) represents the probability that the learner cannot correctly respond to the \( j \)th item with their ability level \( \theta \), and \( Q_j(\theta) = 1 - P_j(\theta) \). To select the most suitable item for a learner, the following information function can be used:

\[
I(\theta) = \frac{1.7^2}{[e^{1.7(\theta - b_j)}][1 + e^{-1.7(\theta - b_j)}]}, \tag{3}
\]

As learner ability \( \theta \) and item difficulty \( b \) come closer, the information function value becomes higher; therefore, an item with a maximum information function value given a learner with ability \( \theta \) has the highest recommendation priority.

IRT is usually applied in the Computerized Adaptive Test (CAT) domain. Chen, Lee, & Chen (2005) substituted learning materials for test items; they designed a personalized e-learning system based on IRT to provide learning paths that can be adapted to various difficulty levels of course materials as well as various learner abilities. The success of using IRT in the CAT and e-learning inspires this research to use IRT to measure learners’ abilities and recommend adaptive learning content to them. A learner’s ability is an important characteristic that should be considered because very difficult content frustrates learners and very easy content bores learners.
Concept sequences and sequential pattern analysis

In a learning process, the learning focus of each phase is called “concept” (Lee, Lee, & Leu, 2009). When learners study domain knowledge, they break their knowledge into small parts, and then rearrange or reorder it into a format that makes sense to them. Learners then develop links between these small concepts until they fully grasp the knowledge (Novak & Gowin, 1984). Epistemological order represents the order of learning concepts (Pólya, 1957). An epistemological order of concepts is called a concept sequence in this research. In order to obtain the concept sequences approved by most users, the UALS use sequential pattern mining to extract the patterns of concept sequences from user-created learning materials on the Web.

Sequential pattern mining is usually used to find a recurring pattern related to time or other sequence. Sequential pattern analysis is usually applied in the business domain to help managers determine which items are bought after other items have been bought (Han & Kamber, 2001), or to analyze the browsing order of Web pages (Spiliopoulou, 2000). This research uses the Generalized Sequential Pattern (GSP) algorithm to implement sequential pattern mining because it is efficient and is able to generate all possible candidate sequences so that missing any actual sequences can be avoided (Srikant & Agrawal, 1995). The output of this algorithm is all maximal sequences in frequent-sequence sets. Maximal sequences are frequent sequences and not included in other sequences. In this study, maximal concept sequences represent the epistemological orders of these concepts, and these orders are accepted by most users. The UALS generates learning paths according to the discovered maximal concept sequences.

A sequential rule is an implication of the form, $X \rightarrow Y$, where $Y$ is a sequence and $X$ is a proper subsequence of $Y$, i.e., the length of $Y$ is greater than that of $X$ (Liu, 2007). This implication means that if a sequence $X$ exists, we can find a sequence $Y$ containing it. In this study, $X$ represents a sequence formed by concepts not understood by a learner, and $Y$ is a frequent concept sequence. If the sequential rule $X \rightarrow Y$ has a high confidence (the proportion of all materials that contain $X$ also contain $Y$), which indicates sequence $Y$ has the relationship of implication with concept sequence $X$; therefore, sequence $Y$ is a potential learning path for the learner.

System architecture

The system architecture is illustrated in Figure 1. The operation procedure can be classified into frontend and backend processes. In the backend process, the system collects materials from Web resources created by Internet users and analyzes the concepts included in them. Such materials and knowledge concepts are recorded in a material database (Step A). On the basis of the material database, the concept analysis process analyzes the concept sequences in materials by sequential pattern mining and stores frequent-concept sequences in a concept-sequence database (Step B and C). In a test items modeling process, an instructor designs pre- and post-tests based on the concepts that were recorded in a concept-sequence database and stores test questions in a test database (Step D and E). Notably, testing is not necessary in an e-learning 2.0 environment; the pre- and post-tests designed here are used for measuring learners’ learning performance and evaluating the system.

The frontend process can be classified into the following stages:
1. Initial stage (Steps 1–3): Learners login this system and select a course to study, and the interface agent requests the learning-path agent to provide learning services.
2. Pre-test stage (Steps 4–7): The learning-path agent notifies the test agent to provide the learner with a pre-test. The test agent analyzes test results to find concepts not understood by the learner and transmits them to the learning-path agent.
3. Path generation stage (Steps 8–9): The learning-path agent applies sequential rules to generate an individual learning path based on concepts not understood by the learner and existent concept sequences in database. When a learning path is generated, a learning-path agent stores it in a user-profile database.
4. Learning stage (Steps 10–17): The learning-path agent notifies a material-recommendation agent to provide a learning content for a given concept. The agent recommends material matching the difficulty level to the learner’s ability. The learner is asked to indicate his/her comprehension level and perceived material difficulty after s/he studies this material. A feedback agent collects learner feedback and re-evaluates learner ability and material difficulty. If the learner is able to comprehend this content, the learning-path agent navigates the next learning concept. Otherwise, the learner continues studying the same concept with different materials. This procedure repeats until all the concepts in the learning path have been learned.
5. Post-test stage (Steps 18~21): If the learning-path agent senses that the learner has already finished the entire learning path, it notifies the test agent to provide post-test questions for the learner.

**Adaptive navigation support**

The system collects user-created teaching materials from Web pages and blogs and divides these materials into concept units. When users share their knowledge on the Web, they usually arrange concept units in order, like a list or a catalog, based on their cognitions about domain knowledge. Each user-created teaching material presents a concept sequence, which represents the user’s notion of learning order. The system discovers the concept sequences that frequently occur in user-created materials. Thus, frequent concept sequences are collaboratively decided by Internet users. The discovered sequential patterns are a kind of collective intelligence and are used to support adaptive navigation.

**Concept sequences analysis**

The following example demonstrates the way to find concept-sequence patterns. Assume that five materials about C++ programming are collected; then, the frequent concept sequences are generated by the following steps:

1. The concept order in each material is presented as a sequence. Table 1 shows the concept sequences in the five materials.

<table>
<thead>
<tr>
<th>Material ID</th>
<th>Concept sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt; data type, overload &gt;</td>
</tr>
<tr>
<td>2</td>
<td>&lt; introduction, data type, process control, class &amp; object, function &gt;</td>
</tr>
<tr>
<td>3</td>
<td>&lt; data type, string &amp; reference, class &amp; object &gt;</td>
</tr>
<tr>
<td>4</td>
<td>&lt; data type, process control, class &amp; object, function, overload &gt;</td>
</tr>
<tr>
<td>5</td>
<td>&lt; overload &gt;</td>
</tr>
</tbody>
</table>

2. Find frequent 1-itemset in which all the items’ support levels are higher than the threshold. In this example, the minimum support is set to 40%, which means that the concept which appears at least twice in the five materials will be a frequent concept.

3. If a concept is not included in frequent 1-itemset, it can be deleted; on the contrary, frequent concepts are mapped to assigned numbers for analysis (see Table 2). Table 3 shows the concept sequences after mapping.

4. Use frequent 1-itemset to find other frequent sequences with different lengths. The GSP algorithm is employed to generate candidate sequences and frequent sequences. Finally, the algorithm returns the maximal sequences. Table 4 shows all frequent sequences and the sequences < 1 5 > and < 1 2 3 4 > are the maximal sequences.
Learning path generation and navigation

To achieve personalization, a learner’s prior knowledge level should be considered to generate the learning path. An instructor predefines the minimum score for each concept in the pre-test to diagnose whether a learner understands these concepts; if a learner does not get scores higher than the predefined scores of some concepts, the system knows that the learner is not able to comprehend these concepts. Notably, this pre-test procedure can be replaced by self-reporting which concepts are not yet comprehended in the e-learning 2.0 environment. Given that the minimum support is 40% and the minimum confidence is 60%, if the pre-test result indicates that the learner cannot comprehend concepts 2, 3, and 5, the learning-path generation by sequential rules is described as follows:

1. According to the maximal concept sequences in Table 4, we can present concepts not understood by the learner as sequences <2 3> and <5>, that are called learner’s un-comprehended concept sequences.
2. Learner’s un-comprehended concept sequences are employed to find their corresponding sequential rules: <2 3> → Y and <5> → Z. Y and Z are sequences that are implied by <2 3> and <5>, respectively.
3. Two rules can be found based on the minimum confidence:
   Rule 1: <2 3> → <1 2 3 4> (sup. = 40%, conf. = 100%). Rule 2: <5> → <1 5> (sup. = 40%, conf. = 67%).
   If a sequential rule, e.g., <2 3> → <1 2 3 4>, does not satisfy the minimum confidence, the sub-sequences of <1 2 3 4>, e.g. <1 2 3> or <2 3 4>, will be considered to find other sequential rules.
4. According to these two rules, the un-comprehended concepts 2, 3, and 5, all have prior concept 1, therefore concept 1 is the first learning concept in the learning path. The confidences of rules will be used to decide the sequence priority because the confidence indicates how strong the rules are. Because Rule 1 has higher confidence than Rule 2, the concept order in the learning path is 1 → 2 → 3 → 5. Notably, concept 4 is not a prior concept of the un-comprehended concepts 2, 3, and 5, and the learner has already understood this concept; therefore it does not need to be included in the learning path.

Adaptive presentation

The adaptive presentation mechanism considers both material difficulty and learner ability because these factors affect the suitability of materials to a learner. The following subsections will describe how to estimate material
difficulty and learner ability, and how to adaptively present materials.

**Adjustment of material difficulty level**

It is inappropriate for an instructor to determine the difficulty levels of course materials because an instructor’s view may not represent learners’ view. This system automatically adjusts the difficulty levels of materials on the basis of collaborative voting approach (Jiang, Tseng, & Lin, 1999; Chen et al., 2005). A 5-point Likert scale is employed to measure learners’ perceptions of material difficulty and the scale ranges from -2 (very easy) to +2 (very hard) to indicate the difficulty levels from \( D_1 \) to \( D_5 \). The difficulty of a concept material is estimated using the following formula:

\[
b_j(\text{voting}) = \frac{\sum_{i=1}^{5} n_{ij} \times D_i + b_j(\text{initial})}{N_j + 1},
\]

where \( b_j(\text{voting}) \) denotes the average difficulty of the \( j^{th} \) concept material after learners give a collaborative vote. The variable \( b_j(\text{initial}) \) is the initial difficulty of the \( j^{th} \) concept material that can be predefined by an instructor or the material provider. The variable \( n_{ij} \) represents the number of learners whose responses belong to the \( i^{th} \) difficulty level for the \( j^{th} \) material, and \( N_j \) is the total number of learners who rate the \( j^{th} \) concept material, and \( N_j = \sum_{i=1}^{5} n_{ij} \).

**Estimation of learner abilities and recommendation of adaptive materials**

Assume that a learner responds to a set of \( n \) materials with response pattern \((u_1, u_2, ..., u_n)\). A response \( u_j = 1 \) means that the learner can understand the selected material \( j \). On the contrary, \( u_j = 0 \) represents that the learner cannot understand this material. Next, Formula 1 is used to calculate the probability that the learner can understand the \( j^{th} \) concept material at an ability level \( \theta \) on the basis of the adjusted material difficulty \( b_j(\text{voting}) \), and Formula 2 is applied to estimate the learner’s ability. The value of \( \theta \) that makes Formula 2 return the maximum value is the new estimated learner ability. Then, the information function shown in Formula 3 is applied to choose the most suitable material for the learner. The concept material with the maximum value of the information function given the new estimated learner ability \( \theta \) is the material that best fits learner ability.

In addition to the adaptive learning-path generation and navigation approach, this adaptive presentation approach is also user-centric. Learners collaboratively rate material difficulties, and their abilities are estimated according to their response patterns and user-determined material difficulties. Both material difficulty and learner ability are adjusted dynamically.

**Experimental design**

This study aims to develop a user-centric adaptive learning system on the basis of collective intelligence. To evaluate its performance, a laboratory experiment was conducted in a computer room. It used a pre- and post-test experimental design to determine the effects of the proposed e-learning system. We chose “webpage design” as the experimental course because its user-created materials on the Web are sufficient to analyze patterns. Moreover, the subjects were undergraduate students who majored in Information Systems, and this course was relative to their field. The test questions in the pre-test and post-test were selected from TQC (Techiciency Quotient Certification) test bank. TQC is a computer literacy certification provided by a non-profit organization in Taiwan. In this experiment, subjects were randomly classified into three groups and requested to study through the course. Different groups of subjects used different learning mechanisms that are described as follows:

Group 1: Traditional e-learning that uses expert-defined learning paths and content—Users study through the learning paths and browse the materials that are pre-defined by an instructor.

Group 2: User-centric and adaptive learning path with expert-defined content—Users study the course following the user-centric learning paths determined by the adaptive navigation mechanism, but the material of each learning node is decided by the instructor.
Group 3: Complete user-centric adaptive learning—Users study the course following the learning paths determined by the adaptive navigation mechanism, and the materials of each learning node are determined by the adaptive presentation mechanism.

These groups utilized the same material resources and concepts that were found by the backend process. The experimental procedure comprises the following phases:

1. Pre-test: Through the pre-test, the system can obtain information about subjects’ prior knowledge levels of the learning concepts.
2. Learning stage: Subjects were randomly classified into three groups and requested to study through the course. Different groups of subjects used different learning mechanisms. Their learning processes lasted for about one hour and unnecessary conversations between subjects were not allowed. Figure 2 illustrates the interface of the learning system.
3. Post-test and questionnaire: Subjects took a post-test when they finished their learning processes. The test results were used to measure their learning performance and the effectiveness of the system.
4. Data analysis: This study analyzes subjects’ learning performance, and estimates learner ability and material difficulty. Data were collected from the pre-test, post-test, and the database.

**Experimental result**

Seventy-nine undergraduate students participated in this experiment; 27 were in Group 1, 26, in Group 2, and 26, in Group 3. The chi-square test for homogeneity reveals that the distributions of genders and learning experiences in these three groups are not different.

**Analysis of learning performance**

To evaluate whether the test questions in the pre-test and post-test are equally difficult, we invited 12 undergraduate students who majored in Information Systems to take these two tests and to indicate their self-perceived difficulties before the experiment. The results revealed that the test questions in the two tests were equally difficult in terms of testing scores and self-perceived difficulties.

Table 5 shows that post-test scores are significantly higher than pre-test scores in the three groups. We further analyzed the difference between subjects’ post-test scores in these three groups. This study applies ANCOVA to treat pre-test scores as a covariate to eliminate its effect on post-test scores. The analysis result revealed that the post-test
scores in the three groups are not significantly different ($F = 1.019, p = 0.366$).

We also analyzed subjects’ self-perceived understanding levels about the domain knowledge before and after learning, using a 5-point Likert scale. The value of this scale ranges from 1 to 5, in which 1 indicates “completely not understood,” and 5 is “completely understood.” The result shows that students in user-centric learning groups (Group 2 and 3) felt that they could comprehend the domain knowledge more clearly after learning, while students in Group 1 did not (see Table 6). Furthermore, we found that increment of understanding level in Group 3 is significantly higher than that in Group 1 ($p < 0.05$, using Tukey post-hoc test).

<table>
<thead>
<tr>
<th>Table 5. Paired-samples $t$ test of pre-test and post-test scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Group 1</td>
</tr>
<tr>
<td>Group 2</td>
</tr>
<tr>
<td>Group 3</td>
</tr>
</tbody>
</table>

*** $p < 0.001$

<table>
<thead>
<tr>
<th>Table 6. Paired-samples $t$ test of self-perceived understanding levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Group 1</td>
</tr>
<tr>
<td>Group 2</td>
</tr>
<tr>
<td>Group 3</td>
</tr>
</tbody>
</table>

** $p < 0.01$

Accordingly, we certify that user-centric learning is comparable to expert-designed learning because they are equally effective. Furthermore, we can infer that user-centric learning can satisfy learners’ expectation and improve learner satisfaction because they can feel the enhancement of understanding level.

Analysis of frequent concept sequences and path lengths

We predefined 34 concepts of HTML (hypertext markup language) domain knowledge according to HTML books and employed these concepts to tag the concept units in online user-created materials. Fifty-three user-created HTML teaching materials were collected from the Web. After analyzing these materials by the GSP algorithm with the minimum support of 40%, 20 frequent concepts (see Table 7) and 125 frequent concept sequences were found.

<table>
<thead>
<tr>
<th>Table 7. Frequent concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

This experiment used frequent concept sequences and sequential rules with a minimum confidence of 50% to construct user-centric learning paths in Groups 2 and 3. To understand the similarity or difference between user-generated and expert-designed learning paths, we list the expert-designed path and user-generated path (generated by sequential pattern analysis) with the 20 concepts in Table 8. These paths illustrate that Internet users and the instructor have four identical notions of concept learning order in HTML domain. They are <3 4>, <2 5>, <10 11 12 13>, and <14 15 16 17 18 19>. For further understanding, some serial concepts are merged as more general concepts if they are relative to each other. Therefore, concepts 3 and 4 are merged as the general “HTML Fundamentals” concept; concepts 11, 12, and 13 are merged as “Hyperlink”; concepts 14, 15, and 16 are merged as
“Table”; and concepts 17 and 18 are merged as “Frame.” Therefore, we find that most Internet users and the instructor have the consensus of the concept learning order: HTML Fundamentals → Layout Tags → Font Tags → Image → Hyperlink → Table → Frame → Form.

<table>
<thead>
<tr>
<th>Table 8. The learning paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>The concepts in learning path</td>
</tr>
<tr>
<td>Expert-designed path 3 → 4 → 2 → 5 → 6 → 1 → 8 → 10 → 11 → 12 → 13 → 7 → 9 → 20 → 14 → 15 → 16 → 17 → 18 → 19</td>
</tr>
<tr>
<td>User-generated path 3 → 4 → 1 → 2 → 5 → 7 → 6 → 8 → 9 → 10 → 11 → 12 → 13 → 14 → 15 → 16 → 17 → 18 → 19 → 20</td>
</tr>
</tbody>
</table>

There are some different notions about the HTML domain between Internet users and the instructor. We list these differences and discuss them in the following:

- **Different notions of Headings concept:** In the user-generated path, users consider that the “Headings” concept should be learned after the general “HTML Fundamentals” concept and before “Layout Tags” concept. In the expert-designed path, “Headings” concept is after “Layout Tags,” “Font Tags,” and “Text Formatting” concepts. Therefore, we infer that users considered that both heading tags (i.e., `<h1>` to `<h6>`) and layout tags (e.g., `<p>`, `<br>`, `<center>`, and `<pre>`) belong to the general “Layout” concept. The possible reason is that heading tags control headlines of text content and they also have a blank line effect like paragraph tag `<p>` and line break tag `<br>`. On the other hand, the instructor considered heading tags as a kind of text formatting tag (e.g., `<b>` and `<i>` because heading tags have bold text effect like the text formatting tag `<b>`.

- **Different notions of Background Setting concept:** In the user-generated path, the “Background Setting” concept (concept 7) is after the “Font” concept. The color setting is an important part in “Font” concept, and the “Background Setting” concept has many settings relating to color (background, text, and hyperlink colors in the page). Therefore, we infer that users consider that color setting is the most important concept of “Background Setting” concept. In the expert-designed path, the “Background Setting” concept is after the learning order “Font Tags → Image → Hyperlink.” Because the “Background Setting” concept has related settings about these three concepts, we can infer that the instructor regarded “Background Setting” concept as a further concept about “Font,” “Image,” and “Hyperlink” concepts.

- **Different notions of Lists concept:** In the user-generated path, “Lists” concept (concept 9) is after the general “Text” concept; therefore, we infer that users considered that list tags (e.g., `<ol>` and `<ul>`) are related to “Text” concept because they are used to arrange a list of text items. In the expert-designed path, “Lists” concept is between “Background Setting” and “Multimedia” concepts. Consequently, we can infer that the instructor regarded “Lists” concept as an individual concept.

- **The different notions of Multimedia concept:** We found that “Multimedia” concept is before the learning order “Table → Frame → Form” in the expert-designed path, but after the learning order in the user-generated path. Generally, a concept sequence is organized in a simple-to-complex sequence for teaching. Accordingly, the instructor considered the concepts “Table,” “Frame,” and “Form” are more difficult for learners than “Multimedia” concept. Additionally, “Multimedia” concept is concerned with embedding audios, animations, or video objects in a Web page, and the instructor considered that audio, animation, and video are other content forms like text and images. Therefore, the instructor preferred teaching “Multimedia” right after “Text” and “Image” concepts. Internet users prefer teaching “Multimedia” concept after “Table,” “Frame,” and “Form” concepts is possibly because Internet users usually write HTML learning materials according to their experience of building websites. Tables, frames, and forms are more pervasive than audio or video in general Web pages. Therefore, Internet users may think “Multimedia” is not the essential concept and put it at the end of learning concepts.

The learning orders of the main concepts in the user-generated path and expert-designed path are similar. However, the users and an instructor may have different notions of some learning orders. An instructor is more concerned with concept difficulty when designing a learning order, whereas Internet users first consider common use and their self-experience when designing a learning order.

The adaptive navigation mechanism generates a learning path according to the frequent concept sequences and the learner’s un-comprehended concepts. The ANOVA analysis shows that the number of learning concepts in Group 1 is significantly higher than those in Groups 2 and 3 (F = 13.953, p < 0.001). This result means that the adaptive navigation mechanism can significantly reduce the lengths of learning paths and improve learning efficiency.
Analysis of learner abilities and material difficulties

To evaluate whether the system is able to estimate learner abilities, we divide the learners in Group 3 into three clusters: learners with high, medium, or low ability. Ability estimation ranges between -3 and +3. That is, the range of low-ability cluster is from -3 to -1, the range of medium-ability cluster is from -1 to +1, and the range of high-ability cluster is from +1 to +3. After eliminating two extreme cases who have an aberrant response pattern (completely understand all materials) in Group 3, 24 subjects remain in the three clusters. We use ANCOVA to analyze post-test scores in the three clusters. The F test reveals that the post-test scores in the three clusters differ significantly ($F = 6.021$, $p < 0.01$). The post-test scores in high-ability cluster are significantly higher than those in the medium-ability cluster ($p < 0.05$) and low-ability cluster ($p < 0.01$). The post-test scores in the medium-ability cluster are almost significantly higher than the post-test scores in the low-ability cluster ($p = 0.067$). The results reveal that the system is able to evaluate learner ability and indicate that using learners’ responses and IRT to evaluate their abilities is suitable in e-learning 2.0 circumstances.

The initial material difficulties were predefined by the instructor. The material difficulties were adjusted through learners’ collaborative voting, and the final material difficulties represent learners’ perceptions of material difficulties. To test whether expert-defined and learner-defined material difficulties are different, we selected 20 materials that have more than 30 learners’ votes and compared their initial and final difficulty estimations. The $t$-test illustrates that a difference exists between the learners’ and the instructor’s perceptions of material difficulties. The final difficulty estimation is significantly higher than the initial difficulty estimation ($t = -2.068$, $p < 0.05$). This result indicates that the instructor underestimated the material difficulties or overestimated the learner abilities.

Conclusions

This study has confirmed the practicability of user-centric adaptive learning. The UALS that employs users’ collective intelligence to generate adaptive learning paths and select materials is comparable to a teaching expert. Students’ learning performances in Group 2 and 3 were not different with that in Group 1. This study also found that learners have more satisfaction and learning efficiency from user-centric adaptive learning. Students in Group 1 did not perceive the improvement of their knowledge levels and the lengths of learning paths in Group 1 were significantly longer than those in Groups 2 and 3. The research results reveal that users may have different notions from an expert in concepts and learning orders; and an expert tends to overestimate learners’ knowledge levels when choosing learning materials. Since a gap exists between experts’ and learners’ views, and learners can easily share their knowledge in the e-learning 2.0 environment, applying collective intelligence to provide formal and direct learning services for the learners is more promising and important in the future learning environment.

Table 9 compares the UALS with typical adaptive learning systems proposed in recent years. The apparent distinction between UALS and existing adaptive learning systems is that learning materials and guidance are collectively provided by users themselves instead of few experts or instructors. This study proposes an effective approach to adaptive learning in e-learning 2.0. However, to adapt learning to learner characteristics, UALS only considers learners’ prior knowledge, abilities, and comprehension levels. Other important characteristics like those of preference, cognitive modality, learning style, and behavior shall be addressed for e-learning 2.0 in the future research. Moreover, evaluating the proposed system and approach in more educational programs is required to clarify its ability of generalization and reliance.

The UALS system totally utilizes user-created content and user-defined learning orders, and therefore, this system enables users to collaboratively create learning services. This system is also able to organize learning services automatically and immediately. The research findings also provide some guidelines to design e-learning 2.0 platforms: (1) An adaptive learning mechanism that can guide users to learn is very important because searching for suitable content and arranging their order by learners themselves is troublesome. (2) A tagging mechanism is necessary to help learners tag knowledge content with the appropriate concepts and supports the mining of concept-sequence patterns. Additionally, taxonomy of concepts is required for generating more general concepts and sequential patterns. (3) A sequence-arrangement mechanism is required to help users arrange and share learning orders based on their cognitions. (4) The knowledge content should be dynamically adjusted for learning based on learners’ perceptions and comprehension levels. Therefore, a feedback mechanism is necessary on e-learning 2.0 platforms. (5) The platforms should record learners’ learning progress to enable learners to decide when to stop or...
resume their learning processes and return their control. Since testing is not necessary in the e-learning 2.0 environment, the platforms should help learners decide which concepts are most appropriate for them to study.

Table 9. Comparison of UALS with existing adaptive learning systems

<table>
<thead>
<tr>
<th></th>
<th>Adaptive Presentation</th>
<th>Material Design</th>
<th>Adaptive Navigation</th>
<th>Learning path Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>UALS (This study)</td>
<td>Based on learners’ abilities and material difficulties collectively determined by users</td>
<td>Created by users</td>
<td>Based on users’ prior knowledge, comprehension levels, and user-generated learning paths</td>
<td>Collaboratively constructed by users</td>
</tr>
<tr>
<td>Reategui’s system (Reategui, Boff, &amp; Campbell, 2008)</td>
<td>Based on learners’ demographic and navigation features and material descriptors</td>
<td>Designed by experts</td>
<td>No specific support</td>
<td>No path model</td>
</tr>
<tr>
<td>Leung’s system (Leung &amp; Li, 2007)</td>
<td>Based on learners’ abilities and learning styles</td>
<td>Designed by experts</td>
<td>Based on learners’ abilities and learning styles</td>
<td>Designed by experts</td>
</tr>
<tr>
<td>TSAL (Tseng, Chu, Hwang, &amp; Tsai, 2007)</td>
<td>Based on learners’ learning achievement and effectiveness</td>
<td>Designed by experts</td>
<td>Based on learners’ learning styles and behaviors</td>
<td>Designed by experts</td>
</tr>
<tr>
<td>Karampiperis’s system (Karampiperis &amp; Sampson, 2005)</td>
<td>No specific support</td>
<td>Designed by experts</td>
<td>Based on learner’s cognitive characteristics and preferences</td>
<td>Designed by experts</td>
</tr>
<tr>
<td>aLFanet (Fuentes et al., 2005)</td>
<td>Based on learners’ knowledge states, learning styles, and cognitive modalities</td>
<td>Designed by experts</td>
<td>Based on learners’ study progress, learning activity, and activities of other learners studying the same subject</td>
<td>Designed by experts</td>
</tr>
<tr>
<td>APeLS (Liu &amp; Yang, 2005)</td>
<td>No specific support</td>
<td>Designed by experts</td>
<td>Based on learners’ abilities, learning time, and difficulty levels</td>
<td>Designed by experts</td>
</tr>
</tbody>
</table>

References


Social Networks-based Adaptive Pairing Strategy for Cooperative Learning

Po-Jen Chuang¹, Ming-Chao Chiang¹*, Chu-Sing Yang² and Chun-Wei Tsai³

¹Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan // ²Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan // ³Department of Applied Geoinformatics, Chia Nan University of Pharmacy & Science, Tainan, Taiwan // flypojen@cse.nsysu.edu.tw

*Corresponding author

(Submitted August 10, 2010; Revised July 11, 2011; Accepted September 6, 2011)

ABSTRACT

In this paper, we propose a grouping strategy to enhance the learning and testing results of students, called Pairing Strategy (PS). The proposed method stems from the need of interactivity and the desire of cooperation in cooperative learning. Based on the social networks of students, PS provides members of the groups to learn from or mimic learners who have good relationships with them. As such, the learning achievement can be raised. Our experimental results show that the proposed method can provide a better learning result, especially when compared to grouping at random, because it enables low achievers to find more appropriate collaborators. More precisely, the experimental results show that the learning achievement of approximately 17.64% of the students is increased. The average score is increased from 79 to 91.9 while the standard deviation is decreased from 16.9 to 13.6.

Keywords

Social learning, Cooperative learning, Pair learning, Grouping strategy, Pair strategy, Learning curve

Introduction

Traditional learning methods have been gradually shifted from individual learning to cooperative learning, because of the ubiquity of e-Learning. By using the fast interaction between people and the obtainment of the test grades, the learning method uses not only the grades, but the selection of partner is also important (Hansen, 2003; Rosenberg, 2001). A good grouping method for cooperative learning has to take into account many factors, such as learning achievement, the depth of teaching materials and categories. As such, personal traits, good relationships, learning behaviors all have something to with the learning results. For the same reason, the friends of an individual can be considered as the incentive of cooperative learning (Mason & Rennie, 2008).

Figure 1 is the framework of friendship network, indicating the attempt to find out appropriate pairing of cooperative learning by the pairing of friendship network. Figure 1(a) illustrates the friendship of each student. If the connection between two students is linear, they are friends. Figure 1(b) shows the member paired within the friendship. Although this study tries to pair the members, the result is not exactly the same as expected. In the example, one of the students in this cluster did not have a friend, and thus, a high achiever was assigned to him.

Pairing strategy (PS) also raised a few questions: (1) Why friendship of individual aid learning? (2) When to use paired learning? From two perspectives, learning cannot be separated from making friends in daily life.
1. Learning cannot be separated from group. In this case, learning is not simply internalizing the course contents. It also includes learning correct knowledge and the methodology to obtain the knowledge. Group provides a place for both observation and mutual verification. Teaching can be considered as explicitness of the internalization of individuals.

2. Course needs interaction. The majority of courses are completed by project or cooperation of several people. No matter which method is used, i.e., by grouping or by relay, completing a course involve different degree of discussion or interaction. For the members, sharing job or discussing the course direction is a kind of direct interaction.

These two statements point out that learning and friends are relevant. They also point out the first step of cooperative learning—collaborators. Apparently, good friends interact most frequently and thus most suitable for collaboration. This explains the second question, i.e., making friends aid learning. As a result, how the cooperative learning works has become a major concern in these researches.

For achievement, we build four modules to achieve this: friendship ranking, achievement detection, dynamic pair assignment, and evaluation. The main contributions of the paper can be summarized as follows:

1. We proposed an efficient pairing strategy for enhancing the learning and testing results of students based on the notion of cooperative learning.
2. The proposed method is built on Hyper Grid Learning System (HGLS http://163.26.2.4/hgls/), an e-Learning system we developed, and is applied to the actual teaching, by pairing based on the seats in the classroom and using it to learn, test, and feedback for both teachers and students.

Related works

Learning is a process of replication and innovation. From the previous experience, the learners can develop their own traits. In this section, we talk about the learning process from modeling to the paired learning.

Social learning

From the perspective of social learning theory, in the social situations, the majority of individual behaviors is learned from observation and modeling. By observing the other individual’s behaviors, an individual can form the concept of how to behave and use this coding information to guide his/her action in the future (Akers, 2009). Bandura et al. (Bandura, 1999) proposed social learning network. They pointed out that learning is not the behaviors of an individual but the results of interacting with surrounding people. These constructed social networks, including interactive behaviors, are called activity theory (Engestrom, 1999). Activity theory is a powerful framework. Its extension more likes psychological construction of constructivism (Fosnot, 1996), social involvement of situated learning (Robey et al., 2000), interaction between individual and environment of distributed cognitions (Hollan et al., 2000). The construction of these networks all emphasizes the interaction between individual and groups so as to accomplish one or more specific objectives.

Friends are found from the structure of socialization. Katz & Kahn (Katz & Kahn, 1966) indicated that, during the process of socialization, the psychological aspect of an individual will play a socialized role and also beings an opportunity to the members to adapt to and synchronize with the society. Krackhardt (Krackhardt, 1990) suggested that, the psychological aspect of cooperative learning focuses on the increase in cognitive ability and the construction of the role as a follower. Johnson (Johnson & Johnson, 1992) pointed out that, the grouping framework design of cooperative learning determines the internal relationship of socialization. Therefore, the relationship of socialization also determines the framework of grouping. Based on the impact of mutual interaction on cooperative learning in (Johnson et al., 2007), it can be inferred that psychological interaction is one of the success factors of cooperative learning.

Cooperative Learning

There are plenty of cooperative learning methods (Johnson & Johnson, 1995; Hancock, 2004). The cooperative methods widely used are: Student Teams-Achievement Divisions (STAD), Team-Game-Tournament (TGT), Jigsaw
II, Group-Investigation (GI), and Co-op Co-op. In terms of the number of learners, it can be divided into group learning and expert learning. The former means that for STAD and TGT, all the students learn together and then practice by group; for GI and Co-op, all the students practice by topics. The latter means that each group has its own topic so that the group leader will try to understand the topic in depth and then extend it to the members of the group. No matter which method is used, it is all based on group teaching and then practice individually. As such, the key for learning is the number of students in each group, and paired cooperation is the easiest way to achieve the requirement. This method shortens the time period of STAD and TGT, and it also contains the expert learning of Jigsaw. The purpose of this learning is twofold: mastery learning and promotion of interpersonal relationships.

Aronson and Blaney (Aronson & Blaney, 1978) indicated that, although it is necessary to consider the grouping method for the initial grouping; after a period of learning, for members with poor learning efficiency, the dynamic grouping adjustment helps reconstruct group learning. As for the group members, some researchers, such as Hooper and Hannafin (Hooper & Hannafin, 1988) found that, the grouping of heterogeneous achievers can lead to better learning achievement. To sum up, both of the grouping strategies can be used at the same time. Johnson and his research group (Johnson et al., 1994; Johnson & Johnson, 1998) suggested that, in grouping algorithm, it is necessary to consider learning achievement and thinking model. These research results indicated that, performance of cooperative learning of the heterogeneous grouping is better than that of the homogeneous one.

As far as e-Learning is concerned, the construction of networks can be moved from the interaction of people to through the Internet. Computer-supported collaborative learning (CSCL) (Hertz-Lazarowitz & Bar-Natan, 2002; Stahl, 2005) is a pedagogical approach wherein learning takes place via social interaction using a computer or through the Internet. By combining the advantages of cooperative learning, using distant learning and local learning via network and computer, learning through the Internet or in the electronic classroom, the cooperative learning can be made more effective (Werner et al., 2004). Because CSCL can not only coordinate the store and retrieval of information, documents, the feedback of teachers and peers but also support group learning, dynamic activity, it can easily support the cooperative learning of traditional learning.

**Paired learning**

Paired learning refers to a learning group composed of two members completing the same learning (VanDeGrift, 2004), in which each one plays a different role (Hanks et al., 2004). In terms of the methods of cooperative learning, paired learning can effectively make students focus on their lessons (Nagappan et al., 2003; Peterson & Miller, 2004). Because the pressure originates from responsibility, students focus more on their share of the work (Williams & Upchurch, 2001). Moreover, they also expect that their achievement can be better (Waite et al., 2004). However, it still has some disadvantages. For example, the two members are more likely to chat with each other (Peterson & Miller, 2004); high achievers may independently complete the assignment originally supposed to be completed by two members (McDowell et al., 2003); the roles of two members may exchange and lead to learning disorder; the difference in opinions of two members (Williams et al., 2002); the lack of time for two members to discuss with each other (VanDeGrift, 2004), and so on. Paired learning minimizes the mimic scope of social learning in the sense that the subject from whom to learn is the collaborator. At the same time, because the number of collaborators is minimized, the interaction is maximized when compared to the other ways of grouping.

**The Problem Definition**

According to Johnson, the number of people for the cooperative learning ranges from 2 to 6 for a simple reason. In other words, 2 is the smallest unit for interaction. The question is: Does there still exist the so-called cooperative nature in the case of cooperation of 2? Friendship takes time to build up. Social learning emphasizes that learning does from observation and modeling. That is why we set 2 as the number of learners in a group for the cooperative learning. The input of this grouping method in this research is the establishment of friendship network and inspection on learning achievement. Questionnaires are used to investigate the friendship network. This friendship network further subdivides the attributes of clusters into: communication network, consulting network, and noxious network. As for learning achievement, it is composed of past tests and is the regression curve indicated on the multidimensional parallel axes for predicting learning trend. As for the grouping, the output, friendship network and
learning trend are combined as the basis. The result of the grouping is tested as the basis for the next grouping. The purpose is to achieve learning. Aimed at the way pair is formed, we propose three hypotheses and verifications:
1. Paired learning is better than individual learning.
2. As far as paired learning is concerned, pairing based on social network is better than grouping at random.
3. As well as the need of low achievers, the grade of high achievers may go down slightly.

The proposed method

Figure 2 shows the proposed grouping strategy, which is composed of four steps: friendship ranking, achievement detection, pairing strategy, and evaluation, in that order. Each box represents the result of a process. First, friends of individuals are classified to construct different friendship network sets for ranking. Then to detect achievement, equation of a regression curve is applied to forecast of possible learning trend and analysis is carried out for past evaluation results to forecast results of next test according to current situations. Finally, strategy pairs up suitable students for learning and verification. For convenience of discussion, evaluation is presented in the section of Experimental Results.

Preliminaries

Friendship ranking

Friendship network is an epitome of social learning network. Friendship network consists of communication network $C$, advisory network $A$, and dislike network $D$. After classification, they can be depicted as shown in Figure 3. That shows the networks of a student, in which the arrow denotes relationship between two students, each shape represents different interaction relationship. In the center, ◇denotes the learner, ◇ denotes the person that the learner likes to communicate to; □ denotes the person learner asks pieces of advice from; △ denotes the person whom the learner dislikes; ○ denotes the person whom the learner has no interaction with.

Figure 3. Networks of a student
Networks relationship between each student and the other students are expressed by a series of vectors, and all students’ vectors form a matrix defined as follows:

**Definition 1. (Relationship Matrix)** For all students \( j \neq k \) and \( j, k \in \{1, 2, \ldots, m\} \), the relationship matrix \( R \) is defined as the relationship vector of \( j \), student \( k \) will communicate with \( j \) and give some pieces of advice by

\[
R_j = [r_{jk} \mid r_{jk} = z_{jk} - d_{jk} - d_{kj}]
\]

where

\[
Z_j = [z_{jk} \mid z_{jk} = t_{jk} \cdot a_{jk}], \text{ is the talk-advice vector of } j \text{ who will ask pieces of advice from } k.
\]

\[
T_j = [t_{jk} \mid t_{jk} = c_{jk} \cdot c_{kj}], \text{ is the dialog vector of } j, \text{ in which two learners will communicate.}
\]

\[
C_j = [c_{jk} \mid c_{jk} = 1 \text{ if } j \text{ communicates to } k; 0 \text{ otherwise.}]
\]

\[
A_j = [a_{jk} \mid a_{jk} = 1 \text{ if } j \text{ asks for pieces of advice from } k; 0 \text{ otherwise.}]
\]

\[
D_j = [d_{jk} \mid d_{jk} = 1 \text{ if } j \text{ dislikes } k; 0 \text{ otherwise.}]
\]

The relationship matrix can take the following values. The maximum \( r_{jk} \) is 1 when \( z_{jk} \) is 1, and \( d_{jk} \) and \( d_{kj} \) are 0 whereas the minimum \( r_{jk} \) is -2 when \( z_{jk} \) is 0, and \( d_{jk} \) and \( d_{kj} \) are 1. From \( R \), the friendship network for each student can be defined as follows:

**Definition 2. (Friendship Set)** The friendship set \( F_j \) of a student \( j \neq k \) is a set composed of four sub-friendship sets \( F_j^I \), \( F_j^{II} \), \( F_j^{III} \) and \( F_j^{IV} \) defined as:

\[
F_j^I = \{ k \mid r_{jk} = 1 \}
\]

\[
F_j^{II} = \{ k \mid r_{jk} = 0 \}
\]

\[
F_j^{III} = \{ k \mid r_{jk} = -1 \}
\]

\[
F_j^{IV} = \{ k \mid r_{jk} = -2 \}
\]

The friendship set, \( F_j^I \), means student \( j \) will communicate with each other and get advices from them among the students, whereas \( F_j^{II} \) means neither students communicate, nor them advice. The friendship sets from \( F_j^I \) to \( F_j^{IV} \) are ranked by the value of relationship matrix. More detailed results of friendship network can be found in the appendix A.

**Achievement detection**

Achievement comes from previous tests. Achievement detection includes two parts, learning behavior-achievement normalization and learning regression functions. Test converts achievements into scores, and the scores indicate the completion rate of learning or learning degree of chapters. The difficulties of different tests are different; however, they can only be compared after normalization. The test grades can be used to draw a regression line, which indicates the learning results of a student over a certain period, and predict the grade of next exam. Our experimental results indicate that \( d = 2 \) or 3 gives the best results (Chuang & Yang, 2008), and the regression curve approaches the actual achievement.

**The pairing strategy algorithm**

The grouping strategy we proposed herein for cooperative learning is called pairing strategy (PS), the aim of which is to find the best partner for each student so as to promote the overall achievements of a class. Basing on the learning curves of students, the PS consists of three steps: First, find the virtual point; then build the quadratic Bézier Curve; and finally select the best partner.

Figure 4 gives a simple example to show the underlying idea of the proposed algorithm, which depicts the expected achievement of an underachiever where two students, \( u \) and \( v \), their regression lines, \( f(u) \) and \( f(v) \), of three tests. According to \( f(v) \), the forecasted achievement \( v \) is not as good as expected; raising score from \( N_3v \) to \( p_3v \) could change \( f(v) \) to \( f'(v) \). The method works as follows: First, two regression segments are put together, which include two past grades and a predicted grade, for instance, \( N_2v, N_3v \), and \( N_4v \), to find the virtual point \( g \). Then, the quadratic Bézier Curve is built, which will pass through point \( p_3v \) at axis \( t_3 \). Finally, the movement is calculated.
Computing center of attraction

The center of attraction is similar to the “center of attraction” in astronomy (Feynman et al., 1963), which attracts the two stars of different weights to make them move towards the center. The movement distance is the inverse ratio of the weights of the two stars. It is the same as attraction between overachiever and underachiever. The attraction point affecting the achievements of the two students is an imaginary attraction point, denoted \( g \), between the achievement points. The point \( g \) is between \( N_{3u} \) and \( N_{3v} \), the following equation is used to find \( g \) position:

\[
\frac{N_{3u}^2 + N_{3v}^2}{N_{3u} + N_{3v}} \quad \text{and} \quad \frac{N_{3u} - g}{N_{3u} - N_{3v}}
\]

the general formulation is as follows:

\[
g = \frac{N_{3u}^2 + N_{3v}^2}{N_{3u} + N_{3v}}
\]

Building Quadratic Bèzier Curve

Bèzier curve is unique in drawing the moving direction of a curve. By changing the ratio of three end points, a straight line can be interpolated as a curve, which has two advantages: (1) the moving direction of the curve can be controlled and (2) the shifting distance of the curve can be calculated. The amount of change can be calculated using the Bèzier Curve. Based on the definition of Bèzier curve, quadratic Bèzier curve (Sederberg, 1997; Sohel et al., 2005) consists of two control points, two ends points, and numerous points. In this case, only a control point can form a curve segment, thus, a quadratic Bèzier curve, defined as follows, is adopted. The control point of quadratic Bèzier curve is replaced by the center of attraction \( g \) attracting two segments to the center. The curve segment of \( f(u) \) is from \( N_{2u} \) to \( N_{4u} \), and curve segment of \( f(v) \) is from \( N_{2v} \) to \( N_{4v} \).

**Definition 3.** (Quadratic Bèzier curve): The Quadratic Bèzier curve is defined as

\[
B(\alpha) = (1-\alpha)^2P_0 + 2\alpha (1-\alpha)P_1 + \alpha^2P_2
\]

where \( \alpha \in [0, 1] \), \( P_0 \) and \( P_2 \) are the end points, and \( P_1 \) is the control point.

The extreme value of \( B(\alpha) \) can be found located at

\[
\alpha_0 = \frac{P_0 - P_1}{P_0 - 2P_1 + P_2},
\]

and its value is

\[
B(\alpha_0) = \frac{P_2P_0 - P_1^2}{P_0 - 2P_1 + P_2}.
\]

From the above, by letting \( P_0 \) denote \( N_{i-1,i} \), \( P_2 \) denote \( N_{i+1,i} \), and \( P_1 \) denote \( g \), the attracted value of achievement of a student \( j \) in the test \( i \) can be presented by

\[
P_g = \frac{N_{i-1,j}N_{i+1,j} - g^2}{N_{i-1,j} - 2g + N_{i+1,j}}
\]
Figure 5 shows the Bèzier curves of overachiever and underachiever based on the above definition of Bèzier curve. The achievement points of student $v$ and $u$, $p_3v$ and $p_3u$, can be determined based on the equation given above and can be used to deduce a new regression function.

Accordingly, the movement distances are $|p_3v - N_3v|$ and $|p_3u - N_3u|$. In other words, the center $g$ moves the achievement of student $v$ from $N_3v$ to $p_3v$, and the achievement of student $u$ from $N_3u$ to $p_3u$.

![Figure 5. The Bèzier curves of overachiever and underachiever](image)

**Selecting Partner**

The partner selection focuses on each student. In pairing, the position of the student and other students are expressed by high and lower scores. The movement, maximum value of attraction change, can be found between high and lower scores as

$$\text{Movement}(u,v) = |N_{3u} - p_{3u}| + |N_{3v} - p_{3v}|.$$  

For all the students $j=1…m$ in one class, the maximum value of the total movement distance needs to be determined and the paired sets are valid. According to the following algorithm, partners can be found through the friendship network, and the total movement distance can be maximized.

**Algorithm: Pairing ($F$, $N$)**

1. **Input:** $F$ is the set of friendship sets and $N$ is a set of scores at test $i$.
2. **Output:** $P$ is the set of best paired teams set; $M$ is the maximum sum of movement.
3. Sort friendship sets $F_\varphi^A$, $\varphi \in \{1, 2, \ldots, m\}$, so that $|F_\varphi^j| > |F_{\varphi+1}^j|$.
4. $P \leftarrow \emptyset$, $M \leftarrow 0$.
5. For each student $\varphi$ do
6.   For each item $F_\varphi^A$, $A = I$ to $IV$, do
7.     For each item $\varphi_h \in F_\varphi^A$, $h < m$, do
8.     Find $\lambda$ where $\lambda \in F_\varphi^A$ and $\lambda \notin P$. The maximum value of Movement($\lambda$, $\varphi$) can be computed as defined in Eq (1)
9.     $M \leftarrow M + \text{Movement} (\lambda, \varphi)$.
10. $P \leftarrow P \cup \{(\varphi, \lambda)\}$.
11. $F \leftarrow F - \{F_\varphi^A, F_\lambda^A\}$.
12. End For
13. End For
14. End For
15. return $P$, $M$

![Figure 6. The Pairing Algorithm](image)
distance with $\phi$, $\lambda$ is not yet paired. Movement is used to calculate the movement distance between the two students $\phi$ and $\lambda$. $p_{a_d} - N_{a_d} + p_{\phi d} - N_{\phi d}$. These paired students become the members in set $P$ in line 10. Line 11 removes the friendship networks of the students $F_{\phi}$ and $F_{\lambda}$ from sets cluster $F$. Line 15 returns all pairing sets $P$, i.e., the best paired group. Meanwhile, achievement is raised to $M$. The pairing result will be verified in next test; if the result is not good, grouping steps are repeated.

The complexity of the proposed method can be easily derived. A student $\phi$ has to find a partner from four friendship sets $F_{\phi}^{I}$, $F_{\phi}^{II}$, $F_{\phi}^{III}$, and $F_{\phi}^{IV}$ of size $n$. In the worst case, each student has at most $n-1$ friends in any one friendship; thus, the time complexity is $O(n^2)$. In other words, the time complexity of cooperative learning is $O(n^2)$. The purpose of pairing algorithm is to classify all the students so as to reduce the amount of computation. That is, if the partner can be found in the first set, then the time complexity can be reduced.

System design and experiment

This experiment is conducted on Fedora 12. Moreover, all the programs are written in PHP 5.5 and the database used in MySQL. The subjects are 68 students from two classes of sixth grade students, with one class as the social pairs and the other class as the random pairs, and their scores are collected throughout the academic year. Moreover, the experiment spans the midterms and finals of grade six and the grouping is conducted twice, once in the first semester while the other in the second semester. Every time, questionnaires are given to rebuild the friendship sets. The experiment uses the grades of the first semester as the pre-test, the grades of the second semester as the post-test, and friendship sets conducted in the first semester. The data of the first semester is not used because students are not particularly familiar with the grouping experiment. Instead, the data of the second semester is used because students can now pay their attention to learning. The student information and scores are input by their teachers. In this paper, all the experiments described in this paper are tested on HGLS, an e-learning system we built as a testbed of the proposed algorithm.

System design

Figure 7 indicates the position of the grouping strategy in the entire system. The system establishes adaptive tests for different paired groups. Because we use the individual grade as the feedback to adapt the friendship set and trigger of re-grouping so as to make the system adaptive. The learning and curricular structures are created based on learning contents, and all the questions in the database are linked with corresponding curricular structure. All the scores of exams can reflect inadequacy of learning. Thus, the tests which can strengthen or supplement the learning will be established. Further verification of results can reflect pairing effect. In terms of the experimental results, the students are repaired. The system steps are as follows:

1. Pre-test: Conduct test before grouping. For example, all 68 students (from two classes) are divided into social group and random group. The random group adopts random pairing, while the social group adopts the method suggested.
2. Conceptual testing analysis: The test involves concept analysis. After the pre-test, the system analyzes scores and displays understanding degree of concept. This often refers to score proportion of certain chapters.
3. Friendship Network Analysis: From questionnaires of students, the friendship networks often change because they cannot obtain help, or no help is offered or relationship falls apart.
4. Re-pairing: Repairing is from the post-test results; the system performs repairing after pairing, post-test and analysis. The pairing strategy is processed in three occasions:
   a. Average score of the classes declines
   b. Score of overachievers declines
   c. No obvious progress
5. Post-test: Post-test is performed after pairing to verify results; the students answer some simple questions, such as learning process and friend relationships after pairing.
6. Translating: Translate learning process; there are three results:
   a. Pass, continue to learn.
   b. Retention, relearn the current chapter.
   c. Relearn, relearn the previous concepts.
According to the trigger of repairing, there are three situations of a class that concern us. First, the average score of the class decline, which indicates that the majority of a class is the low achievers, which in turn indicates that the previous grouping cannot enhance the grade. In this case, we need to reexamine friendship network. Second, the score of overachievers declines. Although grouping may cause such a situation, the degree of decline is small. If it happens together with the partner, it means that the pairing is problematic. Third, the overall grade of the class significantly declines compared to the overall grade last time. If the decline is too large, regrouping is needed because this indicates that the learning situation of the class is generally not good. After the test, the system also sends report to teachers as feedback for adjusting the teaching method and/or materials.

**Experimental results**

Table 1 shows the pre-test and post-test results of the experimental pairs and the random pairs. In the pre-test, the average scores are 79 and 80.8, respectively, and the standard deviations are 16.9 and 16.5, respectively. Two pairs are similar. The post-test results are different. The average scores of the experimental pairs and the random pairs are 91.9 and 87.1, respectively, and the standard deviations are 13.6 and 15.8, respectively. The pairing effect is positive. From the actual data, after pairing 8.82% of overachievers lift their score, and half of underachievers gain better results. However, achievement of 8.82% of overachievers decline. Two box plots of the experimental result is shown in the appendix A.

<table>
<thead>
<tr>
<th>Team</th>
<th>Number of students</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AVG</td>
<td>SD</td>
<td>AVG</td>
</tr>
<tr>
<td>Social pairs</td>
<td>34</td>
<td>79</td>
<td>16.9</td>
</tr>
<tr>
<td>Random pairs</td>
<td>34</td>
<td>80.5</td>
<td>16.5</td>
</tr>
</tbody>
</table>

**Evaluation**

The two-tailed Z test with significant value 0.1 shows that for the pre-test results, the crucial value is 1.96, and the mean difference of the pre-test values is 1.85, indicating that the two teams have no difference. For the post-test results, the crucial value is 1.96, but the mean difference of the two teams is 4.76, which is much greater than the crucial value. Thus, the two teams have significant difference. The significant difference represents experimental results of the two teams using pairing strategy, and also indicates the pairing strategy based on friendship network and test scores are superior to random pairing strategy.
Figure 8 presents the box plot for pre-test and post-test results. The left side is the pre-test results, divided into four range values $V_1$, $V_2$, $V_3$ and $V_4$, and the right side is the post-test results, also divided into four range values. The changes in the number of students reflect the changes. This paper also discusses the changes in the number of students within the four ranges and rise of underachievers and decline of overachievers, totaling six cases.

![Box plot diagram]

Among them, Case 1 represents high achievement value in both pre-test and post-test; it falls on $V_1$; Case 2 represents $V_2$ and case 3 represents $V_3$ in pre-test and post-test; Case 4 represents $V_4$; Case 5 represents the students which achievement declines (Ex., $V_1 \rightarrow V_2$); and case 6 represents the students whose achievement improves (e.g., $V_4 \rightarrow V_2$). Table 2 indicates the number of students in the two teams in pre-test and post-test. The social pairs have more overachievers, and changes in number of students are less than the random group. Better pairing strategy has good effect. The average value of post-test in the social group is higher than that of random group due to more overachievers in the social group. The pairing method is better than the random pairing. It can be inferred that the friendship network can improve learning effect.

<table>
<thead>
<tr>
<th>Case</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social pairs</td>
<td>5</td>
<td>12</td>
<td>0</td>
<td>8</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Random pairs</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

**Analysis**

In social pairs, the results are verified before and after grouping to discuss changes in number of students. The results are shown in Table 3:

- The cooperative learning can result in better learning effects. The average score rises, and standard deviation declines. The experimental results indicate that two pairing methods make achievement better, and that effect of cooperative learning is better than that of individual learning.
- According to verification by two-tailed $Z$, the pairing strategy based on friendship network and testing score is superior to the random paring. Obviously, cooperative learning is superior to individual learning. For pairing strategy, the friendship network can make cooperative learning achieve better effect.
- The results of cooperative learning indicate that most students are in the same range. The rise of score means that most students improve their achievement, so relative position has no change. Before and after pairing, the students are still in the same range.
- In the random group, the score of 14.71% of student increases from $V_2$ to $V_3$, 2.94% of student increases from $V_2$ to $V_4$, and 5.88% of students increase from $V_4$ to $V_2$. In the social group, 8.82% and 5.88% of students increase from $V_2$ or $V_3$ to $V_1$, and 2.94% of students increase from $V_4$ to $V_2$. From these results, cooperative learning can improve achievement. It can be expected that long-time pairing may significantly improve achievement.
- The pairing based on friendship improves achievement, but in the random group, the score of 8.82% of student declines from $V_1$ to $V_2$, and in the social group, 8.82% of students also have the same situation. The results indicate that pairing has the greatest effect on the overachievers who may have lower score, but the average students in $V_2$ and $V_3$ are not affected.
- In the random pairs, 8.82% of students falls from $V_1$ to $V_2$, 2.94% of students moves from $V_2$ to $V_3$. In the social pairs, 8.82% of students falls from $V_1$ to $V_2$. This demonstrates the center of attraction can attract paired students, causing decline of achievement overachievers who spend efforts in helping their partner.
Conclusion

This study proposed a dynamic grouping approach for cooperative learning to improve achievement. The findings indicated that this grouping strategy can help 8.82% of underachievers to raise their score to an average level (i.e., from $V_2$ to $V_1$ and $V_3$) and 8.82% of average students to raise their score to high level (i.e., from $V_2$ and $V_3$ to $V_1$), while 73.53% of students remain the same achievement. This grouping strategy for pairing up students can reduce complexity and time of grouping, and is suitable for the actual class situations. Moreover, communication is simple and independent and dependent variables are easy to random. Friends and learning achievement serve as the condition for grouping, and a suitable learning partner can be found through the friendship network. This strategy can also help students to have a better understanding of lessons and improve their learning achievement. Its effect is as good as the grouping strategy consisting of more than three students. To conclude, the experiment has its limitation. First, to be friends, they must know each other. Thus, it is generally easier to be based on a class or a few classes. Then, there must exist effective communication between students. That is, the dialog between two students must show that they must either agree or disagree to the things they talk about; especially when manipulating the questionnaire, it has to be pointed out. For the same reason, advice must have something to do with problems in learning, and the reason for dislike. Finally, it takes time to do the experiments because friendship needs time to build up so that the cooperation will be as natural as possible. This is probably the only way to verify learning by modeling.

Acknowledgements

The authors would like to thank the editors and the anonymous reviewers for their constructive comments and suggestions on the paper. This research was supported by National Science Council, Taiwan, R.O.C., under Grant 98A063.

References


Appendix A

Social Network Results

Figures 9 to 11 show the process of friendship network, in which 1 to 17 are boys, and the others are girls. The data is based on the results of questionnaire, which asks students if they communicate, ask for advice, and dislike each other. Those present the results of communication, advisory and dislike networks. Figure 12 is the friendship network, which results from the first three figures.

Figure 9. Communication Networks and Statistics

Figure 10. Advice Networks and Statistics

Figure 11. Dislike Networks and Statistics

Figure 12. Friendship Networks and Statistics
Figure 9 reveals the communication network based on the same gender. Figure 10 illustrates that the number of students asking for advice is less than the number of students who communicate to each other. Like communication, the advices mostly happened between the same genders. Figure 11 shows the results of dislike networks; the persons the students dislike are more than those who believe. Moreover, they seem to exclude the opposite sex. For some reasons, the students dislike some people; in other words, the students only like to communicate to certain people. Figure 12 summarizes the three networks and shows that most students fail to form useful friendship. Only a fewer friends can help each other. Obviously, it is difficult to find a partner in the same class.

Experimental Results

Figure 13 shows the results of before and after pairing. The top and bottom squares represent, respectively, the maximum and minimum scores; the upper and lower triangles represent, respectively, the scores of anterior quarter ($Q_1$) and posterior quarter ($Q_3$), and the middle green circle represents the average score.

The two box plots indicate three results:
- The maximum and minimum values of the right box are higher than that of the left box, suggesting that achievement has been improved.
- The quantile difference of the right box is smaller than that of the left side, indicating that scores are concentrated on higher scores after pairing.
- The average score increases. After pairing, the achievement of the whole class has upward trend.
Exploring the Factors Influencing Learning Effectiveness in Digital Game-based Learning

Fu-Hsing Tsai, Kuang-Chao Yu1 and Hsien-Sheng Hsiao1

Teacher Education Center, National Chiayi University, Taiwan // 1Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan // fhtsai@mail.ncyu.edu.tw // kcyu@ntnu.edu.tw // hssiu@ntnu.edu.tw

(Submitted September 6, 2010; Revised March 10, 2011; Accepted May, 29, 2011)

ABSTRACT

This study developed an educational online game, Super Delivery, targeting knowledge about saving electricity, and conducted case studies of eight sixth-grade students using this game to explore the factors influencing the effectiveness of students’ knowledge acquisition in digital game-based learning (DGBL). This study followed Miles and Huberman’s (1994) suggested procedure to analyze qualitative data and find the patterns across cases. Based on the analysis of the case-ordered predictor-outcome matrix and diverse evidence including qualitative and quantitative data, a casual map and a decision tree were constructed to explain why differences in the effectiveness of knowledge acquisition existed among the study participants. It was found that many factors interactively influence students’ effectiveness of knowledge acquisition in DGBL. Students’ learning motivation, learning ability, and playing skill could be key factors that collectively influence the effectiveness of knowledge acquisition in DGBL. Also, students’ learning motivation, learning ability, and playing skill were affected by their playing motivation, prior knowledge, as well as online game experience respectively. The results of this study may help teachers consider how effectively utilizing an educational game for enhancing students’ learning effectiveness in DGBL.

Keywords

Digital game-based learning, Learning effectiveness, Knowledge acquisition

Introduction

Play is a common experience for all human beings, including children and adults. Play is an intense learning experience in which both children and adults voluntarily invest enormous amounts of time, energy and commitment, while at the same time deriving great enjoyment from the experience (Rieber, Smith & Noah, 1998). In the last few decades, a substantial body of research (Csikszentmihalyi, 1990; Provost, 1990; Yawkey & Pellegrini, 1984) in anthropology, psychology, and education has indicated that play is an important mediator for childrens’ and adults’ learning and socialization throughout life. Kerr and Apter (1991) also claimed that play is a suitable and respectable way to describe meaningful learning. However, playing digital game has become the favorite form of play for the ‘digital natives’, Prensky (2001) coined the term to refer to today’s youngsters; in other words, playing digital games is also a meaningful learning activity.

Although the inference above could be opposed by someone who associates digital games with violence or addiction, good digital games, regardless of the level of difficulty, actually have the power to get new players to actively learn. According to Gee (2003), well-designed digital games show us that learning is fun, and there are 36 implicit learning principles embedded in good digital games. Therefore, well-known terms such as educational game, digital game-based learning (DGBL), edutainment, or serious game have been coined to harness the power of digital games for training or education. After years of discussion and research conducted by the proponents of DGBL, a majority of people now believes that digital games promote a learning engagement (Van Eck, 2006). Moreover, with the advent of e-Learning, using digital games for learning is rapidly also becoming a new and popular trend (Aldrich, 2004; Prensky, 2001) for overcoming the hurdle of motivating e-Learners.

Despite DGBL having caught most people’s attention, it does not mean that society is ready to massively adopt DGBL in the educational system (Torrente, Moreno-Ger, Martínez-Ortiz, & Fernández-Manjón, 2009). Recently more and more researchers (Papastergious, 2009; Ke, 2008) have noted that the empirical effectiveness of DGBL is still a mystery. For example, although some studies (Whitehall & McDonald, 1993; McFarlane, Sparrowhawk, & Heald, 2002) indicated that educational computer games have a positive effect in math, science, and military education, several reports (Pierfy, 1977; Randel, Morris, Wetzel, & Whitehall, 1992; Van Eck & Dempsey, 2002)
showed that DGBL is ineffective especially when compared with traditional instruction or traditional e-Learning. Hays (2005) also concluded that there is no evidence indicating that instructional games are the preferred instructional method in all situations after reviewing 274 articles on the design, use, and empirical evaluation of instructional games. Hence, as Gunter, Kenny, and Vick (2008) stated, the effectiveness of an educational game is often based on enhancing learning motivating and social interactions rather than the effectiveness of knowledge acquisition. These circumstances imply that, while DGBL has won the trust of most people in terms of its ability to enhance learning motivation, its power to promote better effectiveness of acquiring knowledge is uncertain. Therefore, it is crucial to explore the mystery of learning effectiveness in DGBL. To solve the mystery, a practical way is to understand the factors influencing the effectiveness of learners’ knowledge acquisition in DGBL. Although many researchers (Mitchell & Savill-Smith, 2004) have indicated that students’ distraction caused by game-playing is the key factor affecting their learning effectiveness in DGBL, this research issue has not yet been much explored.

In view of the above, the purpose of this study was to explore the factors influencing the effectiveness of students’ knowledge acquisition in DGBL by attempting to analyze the reasons why differences in the effectiveness of knowledge acquisition exist among DGBL participants. According to Miles and Huberman (1994), when trying to ascertain why different outcomes exist among cases, the ideal approach is to compare a "limited number of cases." Therefore, a case study of eight sixth-grade students for DGBL was conducted. Also, an educational online game entitled Super Delivery was developed by researchers based on Garris’s game-based learning model (Garris, Ahlers, & Driskell, 2002), for the purpose of teaching knowledge about saving energy and how to calculate the cost of electricity at a primary school level. The main aim of this case study was to investigate the learning effectiveness of knowledge acquisition after playing Super Delivery and to discover the reasons why differences in the effectiveness of knowledge acquisition exist, by analyzing the cases’ characteristics and their learning behaviors.

Method

Participants

This study conducted case studies of eight sixth-grade students using Super Delivery for exploring the reasons that differences in learning effectiveness existed among them. Since the approach of this study attempted to observe few cases’ game playing behavior, purposeful sampling, which is the dominant strategy in qualitative research, was used in the selection of study participants. According to Patton (1990), purposeful sampling seeks information-rich cases that can be studied in depth. In all types of purposeful sampling, the most useful strategy for the naturalistic approach is the maximum variation sampling approach (Lincoln & Guba, 1985), wherein a small sample is selected to reflect maximum diversity across specific attributes for describing significant patterns within the group. Accordingly, for selecting information-rich cases, the purposeful samples of this study were selected by maximizing variations on the dimensions of gender, prior knowledge, and online game experience, from one sixth-grade class of a public primary school in Kaohsiung City, Taiwan. Prior knowledge was defined as the last semester’s mathematics and science grade, since playing Super Delivery need to make use of the arithmetic ability and to learn the science knowledge. Finally, two male students (Kevin and David) and two female students (Jenny and Lisa) with a high level of prior knowledge were recruited, as well as two male students (Allen and Peter) and two female students (Alice and Nancy) with a low level of prior knowledge (see Table 1). However, of the eight study participants, only three students (Jenny, Alice, and Nancy) had little experience playing online games.

Game and instrument

Global warming is a critical issue around the world. One of the practical ways to help prevent the Earth from warming further is to educate the young generation about how to save home electricity use. Based on this idea, this paper constructed a multiplayer educational game targeting five important knowledge concepts appropriate for sixth-grade students: (1) sources of energy; (2) sources of electricity; (3) conductors and insulators; (4) electric power rate; and (5) electricity cost. The game, Super Delivery, was developed by the researchers using the Dimensione X engine. The process of designing the game followed Prensky’s (2001) suggestion about how to combine computer games and learning. First, this paper selected the role playing game (RPG) style and two types of interactive learning
techniques, including task-based learning and drill practice. This was blended with the learning content to create a multiplayer educational game that conformed to Garris’s (2002) repeating game cycle.

The game story hypothetically occurs in the year 2050 when global oil reserves are exhausted. The RPG style puts players in the role of a deliveryman to delivering fast-food on an electric motorcycle in a virtual city. The goal of the game is to complete as many delivery tasks as possible and to compete with other online players. To complete a delivery task, gamers need to find the shortest route in the maze of the virtual environment, and have to overcome two main obstacles. One obstacle is that every electric motorcycle loses power constantly once the motorcycle moves. Thus players have to search for charging stations or equipment such as batteries to continuously supply the electric power of the motorcycle. In the charging station, players need to calculate the correct electricity cost they want to charge. The other obstacle is that a quizzer will interrupt gamers’ movements in specific quiz stations by asking random multiple-choice questions about the five important knowledge concepts mentioned above. The players will be blocked until they provide the correct responses. These two obstacles enhance the difficulty of the game and stimulate players to engage in task-based learning and drill practice. In every delivery task, players can realize the concept of the electric power rate from their moving motorcycles. The electric charging stations also provide opportunities for players to understand how to calculate the cost of electricity. In addition, quiz stations give chances for players to repeatedly learn important knowledge about electricity. Therefore, the game triggers a repeating game and learning cycle, as per Garris’s (2002) game-based learning model.

Super Delivery can be run in any recent major web browser. The interface of this game consists of one window split into three frames (see Figure 1). In the left frame, players can see and control their own avatar’s movement in the game scene, and game messages are displayed automatically when players trigger certain events or objects, as well as personal scores and equipment attained. In the right frame, players can see all the game function buttons, such as giving fast-food to someone, picking up the new equipment, checking the hall of fame, or inputting text to interact with others. The game map for displaying players’ positions in the game also is included in the right frame. However, when a player moves his avatar into a charging station or quiz station, the right frame will also provide an area for inputting the charging cost, displaying random multiple-choice questions, or browsing the learning content about the five important knowledge concepts mentioned above, as shown in Figure 2. In the bottom frame, there is an area for displaying players’ personal messages such as the chat history.

In addition to the basic rules as mentioned above, this game also refers to common rules of commercial RPGs so as to enhance the playability of the game. That is, every player’s avatar possesses attributes like energy, experience, money, or skill. Players can upgrade their characters’ abilities, making them faster to deliver food or more effective at saving electricity, through increasing the numbers of these attributes.
To understand whether Super Delivery promoted learning, a performance test for evaluating the study participants’ effectiveness of knowledge acquisition was necessary. Therefore, based on the learning goals of the game, this paper developed a paper-based performance test. The test was composed of 5 multiple-choice (MC) items and 5 constructed-response (CR) items. The MC items measured the five important factual knowledge mentioned above. The questions used in the MC items were similar to those used in the game’s quiz stations. For example, questions such as “which one is a renewable source of energy?” or “which one is a conductor?” belonged to the MC items. The CR items were math problem-solving exercises that measured whether students could calculate the answers to questions about power rate and electricity cost. For example, “how much will an 800 watt toaster cost to run for two hours if the electric company charges 3 NT dollars for one kilowatt hours?” was one of the CR items. To evaluate the validity and reliability of the performance test, two sixth-grade science teachers from the sampled school vetted all items’ content validity, and the reliability was assessed by internal consistency (Kuder-Richardson 20, KR-20) through a pilot study with 29 sixth-grade students before experiment. A KR-20 reliability coefficient of .60 was obtained. This shows that the performance test has moderate validity and reliability. Besides, the items of the performance test also satisfied the item discrimination and item difficulty criteria in that the average difficulty and discrimination of all items were .59 and .56 respectively through conventional item analysis.

Procedure

The experiment began with a training session, in which all participants were trained to familiarize themselves with the game environment for 40 minutes by playing a Dimensione X game called The Beach. Next, all participants were trained individually to think aloud for two sessions (40 minutes each) while playing The Beach. Since the electricity knowledge in performance test was new for all participants (it can also be proved from the initial reactions when all participants faced the charging station or quiz station in Super Delivery through observations) and in order to eliminate pretest sensitization, no pretest condition was used in this study. Hence, after the training sessions, all participants were asked to play Super Delivery while using the think-aloud method for six sessions (40 minutes each) evenly spread out over six weeks. During the intervention, all participants were asked to sit in a computer lab with no face-to-face interaction, while students could only talk or help each other in game environment. The researchers’ roles were observers. Following the intervention, the performance test was distributed to participants. Finally, each participant was interviewed individually and was asked questions regarding their learning perceptions and electricity knowledge.

Data collection and analysis

The data collection for this case study consisted of both qualitative and quantitative data. The qualitative data were gathered through observations, game-playing records, think-aloud verbal protocols, interviews, and document analysis. Observations, think-aloud verbal protocols and game-playing records were employed when each student played the game. Participants were asked to think aloud to express their cognitive processes and emotional situations when interacting with the game. Researchers sat silently behind the participants and took notes to observe their verbal and nonverbal behavior. The participants’ think-aloud verbal statements were recorded and the game screen was also recorded as audio and video files using screen capture software. The audio files of the think-aloud verbal statements were transcribed and analyzed with the observation notes and game-playing records every week, for the
evidence of the students’ behavior characteristics. The researchers conducted semi-structured interviews when participants finished six game-playing sessions. The interviews gathered information related to the participants’ attitudes toward the game and their practical effectiveness of knowledge acquisition. For example, questions included “Do you think the game is fun?” and “Can you explain what electric power is?” The interviews were recorded using a digital recorder, and were then transcribed and analyzed. Relevant documents (i.e., participants’ feedback sheets after game-playing) were also gathered and analyzed for evidence of the students’ perceptions.

These multiple collection methods provided a means to triangulate the validity of the findings (Eisenhardt, 1989). In the analysis of all relevant qualitative data, this paper followed Miles and Huberman’s (1994) suggested procedure, whereby data reduction, data display, and conclusion verification are interwoven before, during, and after data collection. The data display format chosen for this analysis was a predictor-outcome matrix (Miles & Huberman, 1994).

The quantitative data were gathered through the performance test and participants’ prior knowledge achievements, indicated by last semester's mathematics and science grades. The participants’ previous achievements were gathered at the beginning of the experiment and the performance test was administered at the end of the experiment. All of data were used as evidence to support the analysis of qualitative data.

Results

According to Miles and Huberman (1994), the initial procedure in analyzing qualitative data is data reduction. Thus, this paper first identified the learning effectiveness of knowledge acquisition for each participant. Based on the results of the performance test (a total score of 100 points, M=71, n=8) and interviews, it was found that not everyone acquired perfect knowledge concepts through game-playing. Even most students’ practical effectiveness of knowledge acquisition was not as good as the performance in the performance test (other than Kevin), since most students had misconceptions or had forgotten some important knowledge about electricity during interviews. For example:

Researcher: Have you learned the units of electrical energy from game playing? What is the unit of electrical energy?
David: Joule and watt. (Incorrect)
Researcher: Have you learned what electric power is?
David: It means the electrical energy used for one second. (Correct)
Researcher: What do you think the unit of electric power is?
David: Joule and watt. (Incorrect)

Obviously, David learned the concept of electric power after playing Super Delivery, since his initial reaction was just guessing when he faced the similar question in game’s quiz stations. However, he forgot the unit of electrical energy, since he got the correct answer of the same question in the performance test. In addition to David, Jenny, Lisa and Allen also confused the unit of electrical energy with the unit of electric power. It was also found that Nancy and Peter had misconceptions when calculating the electricity cost on the performance test. According to the results of the performance test and interviews, this paper classified students’ practical effectiveness of knowledge acquisition into three levels – high, middle, and low (see Table 1).

Next, the students’ gaming behavior was analyzed to identify key variables that could serve as predictors of the students’ effectiveness of knowledge acquisition. From the analysis of in-field observations, the students’ game-playing records, feedback sheets, think-aloud verbal protocols and interviews, it was found that most participants were immersed in the game play and thought that Super Delivery was a fun game. For example, in the feedback sheets, Lisa wrote, “I think this game is so exciting, and hope I can play this game at home,” and David commented “I really hope I can play this super fun game all the time.” However, it was also found that some participants paid so much attention to playing the game that they ignored or evaded learning. According to the in-field observations, David, Lisa and Allen were anxious to gain victory. They were frequently concerned with their and their classmates’ game scores when playing Super Delivery, so they also frequently evaded learning. For example, David and Allen “felt that reading learning resources in the charging station is a waste of time,” so they utilized their playing skill to solve the charging problem. Lisa also “felt that reading the learning resources in the quiz station is a waste of time,”
so she solved the quizzes by trial-and-error. Thus, the students’ playing motivation was classified into two levels – very high and high (see Table 1).

In addition, it was found that students had three key gaming behaviors when interacting with the game: (1) learning motivation to seriously learn new knowledge; (2) learning ability to successfully understand new knowledge; and (3) playing skill to solve game problems. For instance, students behaved differently according to learning motivation, learning ability and playing skill in the face solving how to charge their electric motorcycles. When students first attempted to charge their motorcycles in the charging station, Kevin, Jenny, Alice and Nancy immediately read the instructions on how to calculate the electricity cost. Meanwhile, Kevin and Jenny spent a lot of time reading the learning contents until they successfully solved the charging problem, but Alice and Nancy had some difficulty understanding how to charge their motorcycles, even though they also read the learning references for some time. By contrast, David, Allen, Lisa and Peter only attempted to charge by trial-and-error, without reading the learning contents when they first entered the charging station, and immediately left the charging station when they could not charge their motorcycles. After that, David, Allen and Peter never read the learning resources in the charging station. However, despite most students not being able to charge in the charging station, some students attempted to charge their motorcycles through playing skill. David, Allen and Lisa, for example, were good at charging their motorcycles by purchasing a virtual battery, which was more expensive than charging in the charging station. Alice, Nancy and Peter were the only students who did not have the ability to charge their motorcycles through utilizing charging stations or virtual equipment in Super Delivery. They always charged their motorcycles by waiting, because the game will automatically provide a small amount of energy after a player runs out of energy and waits for ten seconds.

<table>
<thead>
<tr>
<th>Table 1. Case-ordered predictor-outcome matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong> (pseudonym)</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Kevin</td>
</tr>
<tr>
<td>David</td>
</tr>
<tr>
<td>Jenny</td>
</tr>
<tr>
<td>Lisa</td>
</tr>
<tr>
<td>Allen</td>
</tr>
<tr>
<td>Alice</td>
</tr>
<tr>
<td>Nancy</td>
</tr>
<tr>
<td>Peter</td>
</tr>
</tbody>
</table>

The evidence mentioned above indicated that Kevin and Jenny displayed high learning motivation and learning ability when they faced the charging problem. Alice and Nancy showed great learning motivation in the face of new knowledge, while they also showed their inability to learn alone and inflexibility to make use of game skill. David, Allen, Lisa and Peter lacked learning motivation when facing game problems that could be solved by reading the learning resources, but David, Allen and Lisa showed alternative ways to solve the charging problem through utilizing their playing skill. Therefore, three key gaming behaviors were indentified in this manner. To reduce the data, this paper classified the students’ learning motivation, learning ability and playing skill into two levels – high and low (see Table 1).

The next step in the analysis process was to construct an information display to handle the data systematically. A case-ordered predictor-outcome matrix was used. This type of display allowed the researchers to maintain individual case integrity and facilitate the task of analyzing the data for trends and patterns across cases (Miles & Huberman, 1994). As shown in Table 1, the data for each case were compiled into a table consisting of the variables under consideration, arrayed according to the predictor (independent) variables and the outcome (dependent) variable. Besides the first column where each case’s name was recorded, columns two to seven consisted of the items identified as predictors to the learning effectiveness of knowledge acquisition.

The final step in the analysis process was to interpret the data in order to discover the reason behind participants’ effectiveness of knowledge acquisition in DGBL. A casual map emerged through pondering the relationship between the predictor variables and the outcome variable in the predictor-outcome matrix. As shown in Figure 3, the participants’ learning motivation, learning ability, and playing skill were the key factors directly influencing whether participants could acquire knowledge from Super Delivery. Also, participants’ learning motivation was negatively
affected by their playing motivation. Participants’ learning ability and playing skill were positively influenced by their prior knowledge and online game experience, respectively. That is, participants’ playing motivation, prior knowledge, and online game experience were the factors indirectly influencing the effectiveness of knowledge acquisition in DGBL.

(1) Learning motivation

The data in the predictor-outcome matrix shown in Table 1 reveals that the students who were classified into the low level of learning motivation could not receive a high level effectiveness of knowledge acquisition in DGBL. Therefore, students’ learning motivation to learn new knowledge in game could positively influence whether they can acquire correct knowledge concepts through game play. An examination of the data for each case found this finding to be reasonable and supported by the evidence. For example, according to the research data, it was found that David, Lisa, Allen and Peter, who were classified into the low level of learning motivation and spent little time reading the educational contents when facing the two main obstacles of charging the motorcycle and answering quiz problems in game, all had misconceptions about the knowledge provided in Super Delivery. Only the students who displayed a high learning motivation (such as Kevin) had a chance to acquire correct knowledge using DGBL.

Moreover, this paper also found that there was a trend that showed that the students who were classified into the very high level of playing motivation did not have a high level of learning motivation to learn new knowledge in the game. For example, David, Lisa and Allen, who were extremely anxious to win the game and were classified into the very high level of playing motivation, only received the low level of learning motivation in Table 1. Their attitudes always viewed reading learning content as a waste of time when competing for the victory of game, which affected their learning motivation. Hence, students’ playing motivation toward a game is the key factor that negatively influences students’ learning motivation to learn new knowledge in a game.

(2) Learning ability

Based on the above inference of the relationship between learning motivation and the effectiveness of knowledge acquisition, it could be asked why cases such as Alice and Nancy, who possessed high learning motivation, could not receive a high level of effectiveness of knowledge acquisition. The answer is because Alice and Nancy had some difficulty comprehending the new knowledge even though they read the learning references in the game for a long time. This shows that the students’ level of learning motivation is not the only factor determining the effectiveness of knowledge acquisition. Students’ learning ability to successfully understand new knowledge in game could also be a factor influencing their effectiveness of knowledge acquisition in DGBL. Other than the examples of Alice and Nancy, this inference was supported by the following evidence.

There is an obvious trend in Table 1 showing that the students who were classified into the high level of learning ability were able to at least receive a middle level of effectiveness of knowledge acquisition even if the students were classified into the low level of learning motivation. For example, despite David and Lisa always solving quizzes by trial-and-error (or guessing) and spending little time to read the learning resources in Super Delivery, they got higher scores in the performance test than all the other study participants except Kevin. This phenomenon represents that David and Lisa’s superior learning abilities, such as induction or problem solving, helped them to get an average
effectiveness of knowledge acquisition, even though they never concentrated on learning during the game. Therefore the ability to learn in a game also positively affected the study participants’ performance in acquiring knowledge.

This paper cannot be certain what kinds of learning abilities will affect students’ performance in acquiring knowledge through DGBL, as the study participants leaning abilities were not measured before the experiment. The components of learning ability that affect students’ performance in acquiring knowledge through DGBL need to be further researched in the future. However, it was found that there is a strong relationship between students’ prior knowledge and their learning ability, as shown in Table 1. Students who were classified into the high level of prior knowledge could receive a high level of learning ability in the game, and vice versa. This infers that students’ previous mathematics and science achievements could be one of the factors influencing students’ learning ability to successfully understand new knowledge in a game. Also, according to the analysis of the quantitative data, a statistically significant ($p < .05$) correlation ($r = .901$) between the participants’ prior knowledge scores and the performance test supports that students’ prior knowledge is one of the positive factors that affect their performance in self-directed learning new knowledge in a game.

(3) Playing skill

The above preliminary results reveal that the students’ behavior in DGBL, including their learning motivation and learning ability toward learning new knowledge, are the only two main factors in determining their effectiveness of knowledge acquisition. However, it was found that Jenny’s effectiveness of knowledge acquisition did not conform to the rule. Although Jenny displayed great learning ability and learning motivation in Super Delivery, she did not acquire correct knowledge through playing the game. After reviewing Jenny’s behavior during the game, both in-field observation and think-aloud protocols indicated that Jenny was not a good game player.

Jenny: Turn right. Turn right. Go straight. I need to go back to the fast-food restaurant as soon as possible….Oh my god. I went the wrong way again. (See the map) Turn around. Turn left. (See the map) All right, this is the right direction…..

As this verbal protocol depicts, Jenny often lost her orientation on the way to the targets of the delivery tasks in the game. She also seldom utilized game skills such as spending the money she earned to purchase equipment, in order to solve the game problems. Thus, her poor playing skill was a possible factor influencing her performance in DGBL. According to the in-field observations, it was found the game rules of delivering fast-food and answering random quizzes repeatedly can promote students to acquire knowledge gradually, no matter if they solve the quizzes by reading the learning contents or guessing. That is, the more students completed the game tasks, the more opportunities they contacted new knowledge. According to observation and the game-playing records, Jenny had fewer chances to face the quiz stations than other players such as Kevin, who also possessed great learning motivation and learning ability. Therefore, Jenny’s learning effectiveness could be affected by her poor playing skill in reducing the opportunities of mastering new knowledge in the game. This finding was also supported by the trend shown in Table 1 indicating that students with a low level of playing skill, such as Alice and Nancy, cannot receive a high level of effectiveness of knowledge acquisition.

Therefore students’ game-playing skill also positively affects their knowledge acquisition in DGBL. Moreover, it was found there is a high relationship between students’ online game experience and playing skill. For example, Jenny, Alice and Nancy, who were classified into the low level of playing skill, seldom played commercial online games. According to the in-field observations and think-aloud verbal protocols, it was found that these students often lost their sense of direction in the virtual environment of Super Delivery and they seldom used the virtual equipment to assist themselves in completing game tasks. The data in the predictor-outcome matrix, as shown in Table 1, also supports that there is a strong relationship between students’ online game experience and playing skill. Hence, students’ previous experience of online games is a key factor that positively affects their skill of playing a new educational online game.

Based on these results, a decision tree was constructed, as shown in Figure 4, to describe why differences in the effectiveness of knowledge acquisition existed among the study participants, and to include likely outcomes of learning effectiveness which did not happen to our participants.
Discussion and conclusion

This study developed an educational online game, Super Delivery, to target knowledge about saving electricity, and conducted case studies of eight sixth-grade students using this game for exploring the learning effectiveness of DGBL. In line with previous researchers’ findings (Papastergious, 2009; Ke, 2008; Gunter, Kenny, & Vick, 2008), preliminarily this study also found that the empirical effectiveness of this educational online game is still not clear, since different effectiveness of knowledge acquisition levels arose among the study participants based on the results of the performance test and in-depth interviews. To ascertain why different learning outcomes existed among the cases, this paper followed Miles and Huberman’s (1994) suggested procedure to analyze qualitative data and to find patterns across the cases. Based on the analysis of the case-ordered predictor-outcome matrix and diverse evidence including qualitative and quantitative data, a causal map was constructed as shown in Figure 3, to explain the factors influencing the study participants’ effectiveness of knowledge acquisition in DGBL. The direct determinant of performance on knowledge acquisition in DGBL could be if a student simultaneously possesses learning motivation to learn new knowledge in the game, learning ability to successfully understand new knowledge in the game, and playing skill to successfully complete the game’s task. Also, it was found that the students’ learning motivation about learning new knowledge in the game was negatively affected by their playing motivation in DGBL; students’ learning ability about successfully understanding new knowledge in the game was positively affected by their prior knowledge; students’ playing skill about successfully completing the game’s task was positively affected by their previous online game experience. Hence, these could be the factors indirectly influencing the effectiveness of knowledge acquisition in DGBL.

Of all the factors, what stood out was that students’ playing motivation affected negatively their learning motivation in the game, which then affected their learning effectiveness. This is consistent with the results of previous studies (Mitchell & Savill-Smith, 2004), which have indicated the common problem of educational games is students’ distraction by game-playing. In this study, some cases just concentrated on completing the game, scoring and winning, and avoided reading the learning content in the charging stations or treated the quiz stations as a guessing game. That is the expected reason why some students could not acquire knowledge through playing. In addition, despite the other key factors derived from this study (e.g., students’ learning ability and playing skill) were discussed infrequently in previous DGBL studies, the reason that these factors affected the effectiveness of students’
knowledge acquisition in DGBL is not difficult to comprehend. This may due to that DGBL is one kind of e-Learning activities. These findings tend to be in agreement with the arguments that point out effective e-Learning requires cognitive abilities such as articulation, self regulation, and self evaluation (McLoughlin & Marshall, 2000), or is influenced by prior computer experience or computer skill (Hong, 2002; Pituch & Lee, 2006). Therefore, the findings of this study seem reasonable.

Several implications can be drawn from this study. First, based on the empirical research, this study proposed a casual map for a better understanding of effective learning factors in DGBL. The casual map has demonstrated that many factors collectively influence the students’ effectiveness of knowledge acquisition in DGBL. The findings of this paper plainly reflect the effectiveness of DGBL is complicated and why many researchers viewed the effectiveness of DGBL as a mystery. Second, according to Prensky (2001), DGBL has been viewed as the games not only just the tools for review and reinforcement but also for primary learning of really hard subject matter. However, the results of this study do not support the claim addressed above. Since if Super Delivery was viewed as the primary learning tool for learning what students do not know as this experiment, it is obvious that Super Delivery was not beneficial to those who do not possess well learning ability or learning motivation based on the results of this research. It can be expected that many participants of this study, such as David, Lisa, Alice, or Nancy, may get a better learning effectiveness from Super Delivery if they can use this game as a learning tool for reinforcement after taking a traditional class. This means that DGBL should be used as a learning tool that supports a traditional class or as a stand-alone e-Learning course based on the complexity of subject matter or students’ characteristics. Third, the findings of this study are useful to those designing or adopting DGBL. This study may suggest that the designers of DGBL should develop an adaptive educational game to meet different students’ demands. The adaptive educational game could be a game with AI which can adaptively adjust the degree of difficulty of the game or provide an intelligent tutor based on students’ demands. From the teachers’ point of view, they could refer to the findings of this study to decide how effective to utilize an educational game. For example, when teachers get an educational game, initially they should consider students’ characteristics based on the complexity of learning contents and refer to the casual map shown in Figure 3 or the decision tree shown in Figure 4. According to the predicting of decision tree, they could determine if using the game as a supplementary leaning tool for a traditional class or as a stand-alone e-Learning course.

Despite the findings of this study derived carefully, improvements can be made in future studies. First, to examine the accuracy of this study, future studies could increase the number of samples by quantitative methods such as structural equation modeling, or conduct more in-field experiments to inspect if the other representative games also lead to the same casual map shown in Figure 3. Second, to derive more useful casual map or decision tree for understanding the effectiveness of DGBL, more factors should be taken into account in the future. As mentioned before, the components of learning ability that affect students’ performance in acquiring knowledge in DGBL need to be explored further. The factors in terms of different game type or different ages of audience also should be considered circumspectly when exploring the factors influencing the effectiveness of DGBL. Finally, a revision of Super Delivery based on the findings of this study can be created and then compared with the original version.

Acknowledgements

The work was partially supported by the National Science Council of the Republic of China under Grants NSC 97-2511-S-003-010-MY3, NSC 98-2511-S-003-024-MY3, NSC 99-2631-S-003-003.

References


The Impact of Adapting Content for Students with Individual Differences

Raymond Flores, Fatih Ari, Fethi A. Inan and Ismahan Arslan-Ari
Educational Instructional Technology, Department of Educational Psychology and Leadership, College of Education, Texas Tech University, USA // rayflores983@gmail.com // fatihari@gmail.com // inanfethi@gmail.com // ismihanarslan@gmail.com

(Submitted October 13, 2010; Revised May 9, 2011; Accepted May 16, 2011)

ABSTRACT

Combining adaptive hypermedia methods with strategies proposed by instructional theory and motivation models, an adaptable tutorial was designed and developed. The aim of this study was to assess whether the goals of an adaptable tutorial, which individualized instruction based on student motivation and prior knowledge, were being met (i.e., knowledge gains and motivation gains) and to identify weak or problematic areas, in terms of usability, where the tutorial could be improved. A total of 186 undergraduate students participated in two stages of this study. The data were obtained through questionnaires, surveys, focus group interviews, and system logs. Preliminary results indicated that overall students were pleased with their experiences however, certain groups of students benefited more than others.

Keywords
Adaptable web-based tutorial, Individual differences, Individualized instruction, Adaptive hypermedia, Online learning

Introduction and background

The effectiveness of online courses has been questioned, especially in relation to meeting the individual needs, perceptions, and learning outcomes of students (Akdemir & Koszalka, 2008; Rovai, 2003). Although web-based environments offer many advantages such as the ability to offer more interactivity, personalized instruction, and more independent learning (Brusilovsky, Sosnovsky, & Yudelson, 2009; Inan, Flores, Ari, & Arslan-Ari, 2010), one of the major challenges of web-based instruction has, and continues to be, accommodating students with differing profiles, expectations, prior experiences, and learning abilities (Abidi, 2009; Dogan, 2008).

Research on individual differences has found that certain learner characteristics such as field independence (Chen, 2010; Scheiter & Gerjets, 2007), high motivation (Artino, 2008), high self-efficacy (Artino, 2008; Yukselturk & Bulut, 2007), high self-regulation (Azevedo, Moos, Greene, Winters, & Cromley, 2008; Yukselturk & Bulut, 2007), and particular learning styles (Bajraktarevic, Hall, & Fullick, 2003; Graf, Liu, Kinshuk, & Yang, 2009) are more supportive of online learning than are other learner characteristics. However, as with face-to-face classrooms, in order to be effective, online instructors must make accommodations for a large proportion of students, who do not possess these characteristics. In addition, instructors must consider the differing demographics of distance learners. In comparison to face-to-face students, studies have found that distance learners are usually older (Bocchi, Eastman, & Swift, 2004; Moore & Kearsley, 2005), generally work full time (Inan, Yukselturk, & Grant, 2009), and are more likely female (Sullivan, 2001; Halsne & Gatta, 2002). These diverse background characteristics, coupled with cognitive and learning style differences among students, add to the complexities of accommodating for individual differences in web-based learning environments.

Adaptive educational hypermedia systems

Adaptive Educational Hypermedia (AEH) systems have been lauded for their ability to accommodate individual differences in online learning. Through the incorporation of various instructional strategies, resources, assessments, and interfaces, AEH systems individualize instruction (e.g., content, interface, and strategies) and provide users with more personalized experiences (Inan & Grant, 2008). In their simplest form, AEH systems gather user information and preferences (Brusilovsky, 2001; Triantafillou, Pomportsis, & Demetriadi, 2003; Tsianos, Germanakos, Lekaas,
Mourlas, & Samara,. 2009); make inferences based on the collected data; and then employ various adaptive methods to accommodate each individual student (Inan & Grant, 2008; Lee & Park, 2007; Shute & Zapata-Rivera, 2007).

Several adaptive systems have been designed and developed in order to accommodate learner individual differences. Examples of such systems include AHA! (Stash, Cristea, & de Bra, 2006) and INSPIRE (Papanikolaou, Grigoriadou, Kornilakis & Magoula, 2003) which adapts instruction based on student learning styles; ELM-ART II which adapts instruction based on knowledge levels and student preferences (Weber & Specht, 1997); INTERBOOK (Brusilovsky, Eklund, & Schwarz, 1998) which adapts instruction based on knowledge level; and AES-CS (Triantafillou, Pomportsis, & Georgiadou, 2002) which adapts instruction based on student cognitive style.

Unfortunately, even though numerous AEH systems have been designed and developed, one of the major limitations in the literature is the lack of evaluation studies which document evidence of their usability and effectiveness in terms of student performance, motivation, and/or attitudes. More than others, AEH systems strongly require evaluation, due to their inherent usability problems (Höök , 2000; Jameson, 2003; Paramythis, Weibelzahl & Masthoff, 2010). However, research suggests that the lack of evaluation, or reporting of, may be due to the difficulties encountered when evaluating these types of systems (Chin, 2001; Masthoff, 2002; Weibelzahl, 2005). These difficulties include, but are not limited to, properly defining the control group, determining evaluation criteria, and selecting appropriate samples (Weibelzahl, 2005). Even with these difficulties, Gena (2005) asserts that the evaluation of adaptive systems is crucial and should be a common practice, especially formative evaluation.

**Formative evaluation of AEH systems**

Formative evaluation, an iterative process of collecting data and information during development, can be used to assess and improve the usability and effectiveness of AEH systems (Gena, 2005; Velsen, Geest, Rob Klaassen, & Steehouder, 2008). Formative evaluation plays a decisive role in many systematic design models because it serves as a method for quality control while concurrently focusing on cost-effective improvement throughout the product-development cycle -- rather than only at the end, as with summative evaluation (Dick, Carey, & Carey, 2005; Brusilovsky, Karagiannidis, & Sampson, 2004).

Formative evaluations of AEH systems in the literature are limited but include evaluations of INSPIRE (Papanikolaou, et al., 2003), AES-CS (Triantafillou, et al., 2002), ISIS-Tutor (Brusilovsky & Pesin, 1998), PUSH (Höök, 1998) and ELMART (Weber & Specht, 1997). Papanikolaou and colleagues (2003) found that students exhibited positive attitudes towards the content structure and felt that INSPIRE was easy to follow and comprehend. Results further revealed that students preferred to have more control over system functionality. Upon collecting student feedback from evaluations of AES-CS, researchers redesi gned the interface and addressed design weaknesses such as limiting the scrolling feature of content pages, keeping a consistent layout throughout the system, and providing users with more control (Triantafillou et al., 2002). Brusilovsky and Pesin (1998) found that adaptive navigation support helped users reduce their navigation efforts, while Höök (1998) found that PUSH required its users to make fewer decisions which, in turn, resulted in less cognitive load. A formative evaluation of ELMART conducted by Weber and Specht (1997) found that novice learners benefited more from the direct guidance provided than more knowledgeable students.

**Purpose of study**

The purpose of this study is to present the findings from the formative evaluation (field testing) of an adaptable tutorial which was designed and developed by researchers at a large southwestern university in the U.S. This adaptable tutorial was designed to individualize instructional content for students based on two characteristics—student motivation and prior knowledge. The goals for formative evaluation were, first and foremost, to assess whether the goals of the tutorial were being met (i.e., knowledge gains and motivation gains) and, secondly, to identify weak or problematic areas, in terms of usability, where the tutorial could be improved.

Specific research questions were:
- Based on beginning knowledge and motivation levels (high vs. low), do students differ in terms of knowledge gains, time spent, and their overall appraisal of the tutorial?
What were students’ perceptions of the adaptable tutorial in terms of visual design, organization/navigation, content presentation, assessment feedback, and technical issues?

Methods

Description of adaptable tutorial

Combining adaptive hypermedia methods with strategies proposed by instructional theory and motivation models, an adaptable tutorial was designed/developed and iteratively evaluated by researchers. Due to the complexity of adapting to multiple learner characteristics, adaptation for this tutorial was limited to two learner characteristics: student motivation and prior knowledge. Many studies have considered prior knowledge (Brusilovsky, 2003; Chen & Paul, 2003) and motivation (Johns & Woolf, 2006; Song & Keller, 2001) to be two of the most important factors that should influence the design of web-based instruction. Further, due to the complexity of adapting to multiple levels of these factors, adaptation was limited to two discrete levels—high and low. Using combinations of high/low motivation and prior knowledge levels, four user model clusters were defined and used to categorize learners. These clusters were: (1) low motivation and low prior knowledge, (2) low motivation and high prior knowledge, (3) high motivation and low prior knowledge, and (4) high motivation and high prior knowledge.

Content for this tutorial came from an undergraduate statistics course where basic introductory statistics topics are introduced (e.g., theory of probability, sample spaces, and probability rules). In this tutorial, content was chunked and presented in three sections consisting of 4-5 dynamic web pages. After organizing the content, researchers customized the design of content (i.e., explanations, examples, and practices) by instantiating research-based instructional design guidelines recommended for students belonging to each of the four predefined user model clusters. For example, students with low prior knowledge (clusters 1 and 3) received additional explanations, resources, and content (Brusilovsky, Yudelson, & Hsiao, 2009; Kalyuga, 2006; Kenny & Pahl, 2009), had less navigation opportunities (e.g., hiding links) (Brusilovsky et al., 2009; Scheiter & Gerjets, 2007); and received well guided, slowly paced guided practice (Clarke, Ayres, & Sweller, 2005; Kalyuga, 2007; Scheiter, Gerjets, Vollmann, & Catrambone, 2009). On the other hand, students with high prior knowledge (clusters 2 and 4), for instance, had less graphics (Song & Keller, 2001); had less structured instruction (Clark et al., 2005; Kalyuga, 2007); had more control over navigation (Deubel, 2003; Kalyuga, 2007; Chrysostomou, Chen, & Xiaohui, 2009) and received minimal guidance or no guided practices (Tuovinen & Sweller, 1999). The design of content for each cluster was then further individualized based on student motivation levels. To support the adaptable tutorial, two databases were designed and developed. One database stored information on the individualized content (e.g., text content, text examples, images, Flash files, etc.) which was indexed with a composite factor value (1, 2, 3, or 4). The second database was used to store student information (i.e., demographics, and motivation and prior knowledge levels).

Upon logging into the tutorial, students were presented a prior knowledge test on probability and a survey which measured their initial motivation. Based on the results of these assessments, students were placed into one of four predefined user model clusters for the section one. As students proceeded, the adaptable tutorial would loop through the database, extract all content with composite factor values associated with the student’s user model cluster, and display that content to the student along with other adaptation techniques specific to their user model cluster. Before beginning each section, students’ knowledge and motivation were reassessed to update their user model cluster for the upcoming section. Based on the results of these assessments, students either maintained their current user model cluster or they were placed into another user model cluster for the upcoming section of the adaptable tutorial. If the students switched to another user model cluster, the adaptable tutorial presented the content along with appropriate adaptation techniques specific to that user model cluster.

Evaluation design and procedures

For field testing purposes, the researchers of this study used a tutor alone evaluation design (Woolf, 2009). In this evaluation design, no control group is used. Rather, only one group of students works with the same tutorial design and the study measures specific outcomes from that group. The goal of this evaluation design is to establish or identify something about the learner which can be used to predict learning outcomes on posttests (Woolf, 2009). For this study, the outcomes of interest were changes in student motivation and knowledge levels. Field-testing of the
adaptable tutorial included two stages. In the first stage, data were collected from participants who studied the tutorial online. During this stage, students’ prior and post knowledge, prior and post motivation and their thoughts about the tutorial were assessed. In the second stage, students studied the material in a computer laboratory setting. In addition, participants were observed by the researchers to identify any technical problems, participants were asked to fill out an evaluation survey, and focus group interviews were conducted.

Throughout this study, combinations of quantitative and qualitative methods were used to obtain diverse perspectives on the design of the tutorial. Data were obtained through questionnaires, surveys, focus group interviews, and system logs (see Table 1).

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Description</th>
<th>Instruments</th>
<th># Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAGE -1</td>
<td>Online Field Testing</td>
<td>- Student Questionnaire</td>
<td>153 Undergraduate students</td>
</tr>
<tr>
<td></td>
<td>Adaptable Tutorial individually at a distance</td>
<td>- System Logs</td>
<td>from various degree programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Motivation Scale</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Achievement Tests</td>
<td></td>
</tr>
<tr>
<td>STAGE -2</td>
<td>In-Class Observed Field Testing</td>
<td>- Observation Form</td>
<td>36 Undergraduate students</td>
</tr>
<tr>
<td></td>
<td>Students used adaptable tutorial in class while</td>
<td>- Formative Evaluation Survey</td>
<td>from various degree programs</td>
</tr>
<tr>
<td></td>
<td>researcher observe/log students’ actions</td>
<td>- Student Questionnaire</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- System Logs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Motivation Scale</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Achievement Tests</td>
<td></td>
</tr>
</tbody>
</table>

**Participants**

The target audience for the tutorial was undergraduate students. For field testing purposes, a total of 186 undergraduates from a large southwestern university participated in two stages. Participants came from six sections of an undergraduate introductory technology course. More than half of the students participating in the initial stage were female (53.3%). Their ages ranged from 18 to 30 years old with a mean age of 19.8. The majority of students participating in the second stage were female (55.6%), between the ages of 17 and 26. Each student received extra credit for their participation.

**Learner characteristics measured**

In order to make inferences as to which predefined user model cluster a student belonged to, researchers had to decide how learner’ motivation and knowledge would be measured. Keller’s Instructional Material Motivational Survey (IMMS) instrument was used to measure student motivation. This instrument is based on the Attention, Relevance, Confidence and Satisfaction (ARCS) motivational design model (Keller, 1987a, b) and is widely applied to the motivational evaluation of computer-based instructional materials (Huang, Diefes Dux, & Imbrie, 2006; Huang, 2010; Inan et al., 2010; Song & Keller, 2001). The IMMS scale consists of five point-likert scale items whose values ranged from 1=Strongly Disagree to 5=Strongly Agree. For each tutorial’ section, students whose average motivation, as measured by the adapted IMMS, was less than or equal to 3 were placed into low motivation groups (clusters 1 or 2 depend on prior knowledge), while those whose average motivation was greater than 3 were placed into high motivation groups (clusters 2 or 4 dependent on prior knowledge levels).

For the assessment of knowledge, a locally developed 10 item multiple choice instrument was used. Items measured prior knowledge for each section of the tutorial. Students were given 1 point for correct responses and 0 points for incorrect responses. In addition, one item from each section was weighted by two points for its difficulty. For each section of the tutorial, students scoring less or than equal to 50 percent of the available points for that section were placed into the low knowledge group (clusters 1 or 3 dependent on motivation levels), while those scoring higher than 50 percent of the available points for that section were placed into the high knowledge group (clusters 2 or 4 dependent on motivation levels).
Field testing instruments

In the field-testing phases of this study, various data collection techniques and instruments were used for assessment and evaluation purposes. These include a(n):

- **Student Questionnaire**: Instrument used to collect student perceptions about the adaptable tutorial. To obtain student appraisal, participants were asked to rate the following question, “What is your overall rating of the tutorial?” The response scale ranged from 1 (Poor) to 5 (Excellent). In addition, this questionnaire gathered students’ opinions about what they liked most and least about the adaptable tutorial.

- **Evaluation Survey**: Survey used to gather students’ perceptions on different utilities provided within the tutorial. It included five subsections which were visual design, organization/navigation, content presentation, assessment feedback, and perceptions (Elissavet, & Economides, 2003; Kay & Knaack, 2009; Nordhoff, 2002; Sahari, Abdul Ghani, Selamat, & Yunus, 2009). Students were requested to mark a 5-point scale ranging from “Strongly Disagree” (associated with score 1) to “Strongly Agree” (associated with score 5) for each statement.

- **Observation Form**: Form developed to help the researchers take notes of any technical issues encountered while participants studied the materials.

- **System logs**: These were used to track the amount of time students spent on each page and total completion time.

- **Interview Guide**: A guide which was used to list interview questions and outline the topics to be investigated. Interviews were conducted to get in depth information about students’ experiences with the tutorial.

Data analysis

In this study, several factorial and mixed-design ANOVAs were used to investigate how different clusters of students benefited from the tutorial in terms of knowledge gain, time spent, and appraisals of the tutorial. In addition, descriptive statistical techniques were used to assess students’ thoughts on the tutorial. Qualitative data from interviews, open ended questions, and class observations were collected to supplement quantitative results. Following the collection of qualitative data, data were transferred to an electronic format and then analyzed through iterative cycles of data examination, exploration of similarities and differences among the participants, and a search for confirming and disconfirming evidence that could be incorporated into the conclusions (Merriam, 1998; Miles & Huberman, 1984).

Results

Knowledge gain group differences

The goal of the first research question was to investigate whether certain groups of students benefited more from the tutorial in terms of knowledge gains, than others. To answer this question, researchers used two mixed design ANOVAs. Using beginning prior knowledge as a between subjects factor and knowledge gains as a within-subjects factor, the first examined which group of students, high versus low beginning prior knowledge, benefited more from the tutorial in terms of knowledge gains. Results revealed a significant interaction between beginning prior knowledge groups and knowledge gain from pre to post test, $F(1,184)=12.97, p<0.001$. This indicated that the two groups benefited differently from the tutorial. Researchers then conducted separate paired sample t-test to examine whether this knowledge gain was significant for each group. Results showed that students in the low beginning prior knowledge group significantly increased their post knowledge test scores, $t(146)=-5.82, p<0.001$. On the other hand, there was no significant knowledge gain for students with high beginning prior knowledge. The second mixed design ANOVA examined which group of students, high versus low beginning motivation level, benefited more from the tutorial in terms of knowledge gains using beginning motivation level as a between subjects factor and knowledge gains as a within-subjects factor. Results from this analysis indicated that student knowledge significantly increased after studying the tutorial ($F(1,184) = 9.05, p<.05$) for both groups. However, the interaction between beginning motivation level and knowledge gain was not significant indicating that both motivation groups benefited similarly from the tutorial.
Time spent on tutorial group differences

The second goal was to determine which groups of students spent more time on the tutorial. Two separate ANOVAs were conducted to examine time spent on the tutorial considering beginning motivation level and prior knowledge as between subject factors. Results from the first analysis found that students with high beginning motivation levels spent more time on the adaptable tutorial ($M=18.06$, $SD=10.51$) than the students with low beginning motivation ($M=13.64$, $SD=9.97$), $F(1,182)=7.95$, $p<.05$. The second analysis, revealed that students in high beginning prior knowledge group spent more time studying the tutorial ($M=19.16$, $SD=10.71$), than those belonging to the low knowledge group ($M=15.61$, $SD=10.36$). However, this difference was not statistically significant.

Tutorial appraisal group differences

The third goal was to determine which groups of students were satisfied more with the tutorial. Two separate ANOVAs were conducted to examine students’ overall appraisal of the tutorial considering beginning motivation and prior knowledge levels as between subject factors. In terms of appraisal ratings, no significant difference was found between students with low and high motivation levels. However, the results from the analysis indicated that students from the high beginning prior knowledge group rated the tutorial significantly higher than those from the low beginning prior knowledge group, $F(1,182)=6.89$, $p<.05$.

Students’ perceptions of the tutorial

The second goal of the evaluation was to identify weak or problematic areas in terms of usability. Various data sources including a student questionnaire, an evaluation survey and focus group interview were combined to evaluate students’ perception of the system. Data were collected on student perceptions related to visual design, organization and navigation, content presentation, assessment and feedback, and technical issues.

According to the survey results, participants were overall satisfied with the visual design of the system ($M=3.94$, $SD=0.39$), organization/navigation of the system ($M=4.11$, $SD=0.49$), the presented content and examples provided in the system ($M=3.62$, $SD=0.44$) and the assessment/feedback of the system ($M=3.50$, $SD=0.97$). Moreover, participants’ perceptions about the use of multimedia in the tutorial were moderately high ($M=3.43$, $SD=0.49$) on a five-point Likert scale. These results were also supported with the interview and student surveys. Table 2 summarizes these findings.

<table>
<thead>
<tr>
<th>Most liked</th>
<th>Least Liked</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visual design</strong></td>
<td>Readable text, pleasing graphics and animations, and high quality animations</td>
<td>Plain and dull colors</td>
</tr>
<tr>
<td><strong>Organization/navigation</strong></td>
<td>Well-organized, easy to follow, easy and simple navigation and quickly loaded pages</td>
<td>Scroll down structure</td>
</tr>
<tr>
<td><strong>Content presentation</strong></td>
<td>Segmented simple pages, useful examples and practices, and supportive graphics and photos</td>
<td>Not beneficial for people with high prior knowledge</td>
</tr>
<tr>
<td><strong>Assessment/feedback</strong></td>
<td>Beneficial practices and examples</td>
<td>Too many knowledge assessments and some difficult assessment questions</td>
</tr>
</tbody>
</table>

Table 2. Summary of student evaluation of the tutorial.

During the second stage of formative evaluation, students were observed while they studied the tutorial in a computer lab setting. A few computer screens were observed to freeze while the students were studying, but these technical difficulties were minimal and isolated.
Summary of major findings

The summaries of key findings are presented below:

- Low prior knowledge students benefited more from the tutorial, in terms of knowledge gains, than high prior knowledge students.
- Both high and low motivation groups of students benefited similarly in terms of motivation gains.
- Students with high motivation spent more time studying the tutorial than low motivation students.
- Overall, students found the tutorial easy to read, comprehend, navigate, and enjoyed interactivity. However, some students felt that the topic was boring, content was not relevant to their majors, and did not like the long surveys and assessment questions.

Discussion

For this study, researchers conducted two stages of formative evaluation on an adaptable tutorial prototype with undergraduate students with various backgrounds. Results from the formative evaluation found that low prior knowledge students seemed to benefit more from the tutorial, in terms of knowledge gains, than students of high prior knowledge. These results were similar to results from ELMART (Weber & Specht, 1997) which found that novice students benefited more from adaptive support provided by the tutorial than expert learners. In terms of motivation, both high and low motivation groups benefited similarly from the tutorial. Results from appraisal ratings indicated that overall students were satisfied with the adaptive tutorial. Similar indications of positive student satisfaction were found from the evaluation of INSPIRE (Papanikolaou et al., 2003), AES-CS (Triantafillou et al., 2002), and ELMART (Weber & Specht, 1997).

In terms of tutorial usage, results indicated that high motivation students spent more time studying the tutorial than low motivation students. These results are consistent with Hodges (2004) who asserts that motivated learners are more inclined to continue learning. Some students indicated that the topic was boring and not relevant to their major. Research has indicated that by implementing relevance enhancing strategies, student motivation and performance will increase, especially for the learners who do not find the material interesting or relevant (Artino, 2008; Song & Keller, 2001).

Analysis of interview data and open ended questions found that students felt the tutorial was easy to read and follow, enjoyed the interactivity, and liked the navigation and organization. It was also discovered that some of the students felt that the material was too easy, while others found assessment questions too hard and difficult to understand. Many students reported dissatisfaction with the intermediate assessment surveys which they felt were “long and repetitive.” As from the evaluation of AES-CS (Triantafillou et al., 2002), students from this study suggested that more information be provided to them about the number of sections, questions, and anticipated completion time.

Conclusion

AEH systems provide many advantages to students because they more closely address the issue of individual differences. However, the impact of such systems may be limited if only certain groups benefit more than others and/or if usability issues become problematic. Because the design and development of AEH systems is time consuming and costly, formative evaluation of such systems is crucial and should become a common practice (Brusilovsky et al., 2004; Gena, 2005; Paramythis, Weibelzahl, & Masthoff, 2010). In this study, formative evaluation results encouraged researchers to reconsider certain aspects of their tutorial’s current design and to offer recommendations for future research.

First, although the intended goal of this tutorial was to benefit students with differing knowledge and motivation, results indicate that low prior knowledge students benefited more than high prior knowledge students. It is therefore recommended that future research explore how the design of content can be adapted to better meet the needs of more advanced learners. Instructional design strategies which have been shown to be effective for advanced learners include: less structured instruction (Clarke et al., 2005; Kalyuga, 2007); multiple links to additional resources
(Brusilovsky et al., 2009), more control of navigation (Chrysostomou et al., 2009; Deubel, 2003; Kalyuga, 2007), and faster transition between topics (Reisslein, 2005).

Secondly, while the collection of student data was vital for placing students into their respective user model clusters, evaluation results indicated that one of the most disliked features of the tutorial was the frequency and number of assessment items. In addition to the negative impact on student perceptions, previous research suggests that the intrusiveness of assessments may have unintended effects on student motivation and performance (Salden, Aleven, Schwonke, & Renkl, 2010). It is therefore recommended that future research investigate alternative evaluation structures and/or techniques for collecting data from students learning from adaptable tutorials. Some of these assessment techniques which show promise include the use of Rapid Dynamic Assessment (Kalyuga & Sweller, 2004, 2005) and/or Bayesian Knowledge Tracing algorithm (Corbett & Anderson, 1995).

Third, although measures were taken to maintain and/or enhance student motivation with the design of content, evaluation results indicate that some students describe the topic as boring. Future research on adaptable tutorial which individualize instruction based on student motivation, may consider embedding more relevance enhancing techniques, such as including authentic content which relates to students’ majors, interests, and/or hobbies. Research has found that relevance is one of the most important factors to consider when designing motivational instruction (Edelson & Joseph, 2004; Kember, Ho, & Hong, 2008).

Finally, although this study examines important issues that should be considered when designing an adaptable tutorial (i.e., examination of whether the tutorial allows students to meet objectives and the identification problematic usability issues) (Höök, 2000; Woolf, 2009), future studies should attempt to more closely examine whether the benefits of such tutorial or systems are due to adaptation (e.g., successful user modeling and adaptive decision making) or to the design of the system in general (e.g., system interface). Previous studies have found that one of the major challenges of evaluating adaptive systems is the difficulty of isolating the “adaptivity” component of the system. However, recent adaptive system evaluation frameworks, such as those proposed by Paramythis et al. (2010), show promise by providing a layered evaluation approach, in which, adaptivity is “decomposed” and evaluation is conducted in a “piece-wise” manner.

References


Rovai, A. (2003). In search of higher persistence rates in distance education online programs. *Internet and Higher Education, 6*(1), 1-16.


An Ecological Approach to Learning Dynamics

Peeter Normak, Kai Pata and Mauri Kaipainen1

Tallinn University, Institute of Informatics, Narva mnt 25, 10120 Tallinn, Estonia // 1Södertörn University, Department of Communication, Media and IT, 14189 Huddinge, Sweden // peeter.normak@tlu.ee // kai.pata@tlu.ee // mauri.kaipainen@sh.se

(Submitted November 3, 2010; Revised June 8, 2011; Accepted August 9, 2011)

ABSTRACT

New approaches to emergent learner-directed learning design can be strengthened with a theoretical framework that considers learning as a dynamic process. We propose an approach that models a learning process using a set of spatial concepts: learning space, position of a learner, niche, perspective, step, path, direction of a step and step gradient. A learning process is presented as a path within a niche (or between niches) in a learning space, which consists of a certain number of steps leading the learner from the initial position to a target position in the dynamically changing learning space. When deciding on steps, the learner can take guidance from learning paths that are effective from a viewpoint of the learning community.

Keywords

Ecological approach, Learning space, Affordance, Learning modeling

Introduction

Both the means as well as the requirements of an educational process, have tremendously changed during the last few decades due to the explosion of educational software. There is a shift from the teacher-controlled approach towards a learner-directed approach in planning learning goals, learning environments composition, and learning resources (Attwell, 2007; Brown & Adler, 2008; Anderson, 2009). Due to the increased availability on the Web of teaching and learning applications, traditional learning design models have been criticized (Underwood & Banyard, 2008; Fiedler & Pata, 2009; Väljataga & Laanpere, 2010). Even the nature of education itself is changing; the prioritization of informal learning experiences, besides formal education, expands the range of learning options. Contemporary learning designs need, therefore, to focus on selective combinations of technologies (Bower, 2008). A variety of flexible, self-combinable web-based learning/teaching tools and social software has appeared that may be applied as personal learning environments (PLE) as well as collaborative knowledge building environments. Furthermore, in such environments learning communities are dynamically formed that actively shape their members’ teaching and learning potentials. The conditions of learning have changed also. Instead of it being mainly an individual effort in a clear-cut and teacher-defined learning space along the same path shared by all students, learning has moved towards being simultaneously autonomous and collaborative, taking place in a dynamically changing environment. Whether in formal or informal settings, learners can follow their personal learning paths while being simultaneously guided by the community of learners who collectively shape and change the learning settings. However, a universal theoretical framework for the adequate analysis and modelling of such learning processes in dynamically evolving environments is still missing.

In this paper we propose an ecological approach to learning processes. According to this, learning takes place in a dynamically evolving learning space that is formed not only by the individual learner, but also to a great extent by the wider community of learners and teachers. In order to explain the ecological approach in the learning process, we present its underlying principles and analyse them in relation to existing learning design frameworks. To formalize our approach, we define the basic concepts and propose a general framework for constructing learning paths. The framework is illustrated by two examples: firstly, a simple one for explaining the basic concepts (learning addition of natural numbers) and secondly, a case study for demonstrating the application of the dynamical ontospace model, to guide the learning process. Finally, we outline the basic preconditions of applying the ecological approach for learning design, but only on a formal level at this stage. In order to apply it in practice for analyzing and designing learning processes, more research and development is needed.

Underlying Principles

Classically, it has been assumed that learning is sequential, and the decisions with respect to learning objectives, activities and environments are largely teacher-determined. However, the classical system approach to learning does

An Ecological Approach to Learning Dynamics

Peeter Normak, Kai Pata and Mauri Kaipainen1

Tallinn University, Institute of Informatics, Narva mnt 25, 10120 Tallinn, Estonia // 1Södertörn University, Department of Communication, Media and IT, 14189 Huddinge, Sweden // peeter.normak@tlu.ee // kai.pata@tlu.ee // mauri.kaipainen@sh.se

(Submitted November 3, 2010; Revised June 8, 2011; Accepted August 9, 2011)

ABSTRACT

New approaches to emergent learner-directed learning design can be strengthened with a theoretical framework that considers learning as a dynamic process. We propose an approach that models a learning process using a set of spatial concepts: learning space, position of a learner, niche, perspective, step, path, direction of a step and step gradient. A learning process is presented as a path within a niche (or between niches) in a learning space, which consists of a certain number of steps leading the learner from the initial position to a target position in the dynamically changing learning space. When deciding on steps, the learner can take guidance from learning paths that are effective from a viewpoint of the learning community.

Keywords

Ecological approach, Learning space, Affordance, Learning modeling

Introduction

Both the means as well as the requirements of an educational process, have tremendously changed during the last few decades due to the explosion of educational software. There is a shift from the teacher-controlled approach towards a learner-directed approach in planning learning goals, learning environments composition, and learning resources (Attwell, 2007; Brown & Adler, 2008; Anderson, 2009). Due to the increased availability on the Web of teaching and learning applications, traditional learning design models have been criticized (Underwood & Banyard, 2008; Fiedler & Pata, 2009; Väljataga & Laanpere, 2010). Even the nature of education itself is changing; the prioritization of informal learning experiences, besides formal education, expands the range of learning options. Contemporary learning designs need, therefore, to focus on selective combinations of technologies (Bower, 2008). A variety of flexible, self-combinable web-based learning/teaching tools and social software has appeared that may be applied as personal learning environments (PLE) as well as collaborative knowledge building environments. Furthermore, in such environments learning communities are dynamically formed that actively shape their members’ teaching and learning potentials. The conditions of learning have changed also. Instead of it being mainly an individual effort in a clear-cut and teacher-defined learning space along the same path shared by all students, learning has moved towards being simultaneously autonomous and collaborative, taking place in a dynamically changing environment. Whether in formal or informal settings, learners can follow their personal learning paths while being simultaneously guided by the community of learners who collectively shape and change the learning settings. However, a universal theoretical framework for the adequate analysis and modelling of such learning processes in dynamically evolving environments is still missing.

In this paper we propose an ecological approach to learning processes. According to this, learning takes place in a dynamically evolving learning space that is formed not only by the individual learner, but also to a great extent by the wider community of learners and teachers. In order to explain the ecological approach in the learning process, we present its underlying principles and analyse them in relation to existing learning design frameworks. To formalize our approach, we define the basic concepts and propose a general framework for constructing learning paths. The framework is illustrated by two examples: firstly, a simple one for explaining the basic concepts (learning addition of natural numbers) and secondly, a case study for demonstrating the application of the dynamical ontospace model, to guide the learning process. Finally, we outline the basic preconditions of applying the ecological approach for learning design, but only on a formal level at this stage. In order to apply it in practice for analyzing and designing learning processes, more research and development is needed.

Underlying Principles

Classically, it has been assumed that learning is sequential, and the decisions with respect to learning objectives, activities and environments are largely teacher-determined. However, the classical system approach to learning does
not meet the requirements of self-directed learning in a learning community (Pata, 2009a). The new learning approach that considers self-directed learning in learning communities has at least three major implications for the learning design:

1) Study groups should be viewed as more or less temporary, heterogeneous and less strictly defined in terms of an individual learners’ learning goals, competences, selection of learning paths, and activation of learning resources and tools;
2) In order to satisfy individual learning needs, more autonomy in decision-making and self-direction should be given to the learners while participating in study groups;
3) Learning and collaboration in study groups of self-directed individuals assumes the dynamic emergence and availability of certain well-established rules of behavior in the shared learning space that are defined by the learners themselves, and that can be used for personal or group navigation within the learning space.

To apply the requirements of learner-directed learning design one needs to consider the learning process as an emergent phenomenon. Our approach is influenced by the post-positivist understanding of cognition and human behaviour developed by James Gibson in ecological psychology (Gibson, 1979) and also by the activity theory (Leontiev, 1978; Engeström, 1987; Jonassen, 2000; Conole, 2008). We assume that conceiving the learning process as emergent and dynamic has certain analogies to ecology in the nature, namely, how an individual specimen of any species adapts itself to the niche(s) of its species in the natural ecosystems. In our framework, an individual learner or a group of learners represents an individual specimen, while the community of learners represents one species. Inspired by this analogism, we will apply in this paper the concepts of ecology for describing and designing the learning process.

Gibson’s ecological explanation of the actor’s interaction with the surrounding environment is based on the concept of affordance, defined as a possibility of action dynamically emerging in the environment. This concept has had a significant impact on the approaches to learning design. For example, Kirschner, Strijbos, Kreijns & Beers (2004) suggest an affordance-based and learner-centred sequential interaction design model of learning. Using affordances makes the selection of appropriate tools for certain learning design ecological, since affordances depend on the learners’ perception and action, as well as on the existing possibilities in the learning environment. However, Kirschner et al. (2004) do not particularly emphasize learners’ self-monitoring and self-evaluating activities while performing learning actions. Instead they rather leave the task of monitoring affordances to the teachers. Thus, this design model only partially supports self-directed learning.

Furthermore, Fiedler and Pata (2009) have suggested that the learners should take part in identifying and negotiating the affordances of their individual and joint learning space. Pata (2009a, 2009b) outlined the principles of an ecological learning design framework for supporting self-directed learning in the new social Web. According to this framework, learners and facilitators participate jointly in the construction of a learning environment, modifying it and contributing to the evolution of learning. They can simultaneously use hints acquired from the learning environment and follow the actions of each other.

However, these general aspects of an ecological learning design model do not describe in detail how individual learners would determine their learning paths and navigate in the learning space. As a method of exploration in the learning space and design of learning processes for self-directed learning, the spatio-dynamic ontology approach is proposed in this paper. We adopt a set of spatially defined and ecologically inspired concepts to describe the learning process in such a way that it would enable development of appropriate design models for self-directed learning in dynamic learning environments.

**Basic Concepts**

This section is devoted to the definitions of concepts used in our approach. In order to allow application of the ecological learning approach in different educational settings, the definitions are given in a fairly general terms. These concepts are used in subsequent sections for the formal description of the ecological approach to learning as well as in case studies.
Learner and learning community

The concept of learner denotes an individual or a group of individuals who share a common goal in a learning process – in performing systematic actions for gaining knowledge or improving comprehension or skills. A community of learners (or, equivalently, a learning community) denotes a set of learners who share some common identity. Common identity in turn means that the learners: 1) are engaged in common learning activities, 2) share certain imaginations and 3) align, control and coordinate their actions within the community (Wenger, 1998). The community may have a fixed membership or it may change in time; alignment to such communities is perceived and may be recognized by learners themselves. The learner may simultaneously belong to several communities.

Learning (onto)-space

A learning space is determined by a collection of descriptive dimensions describing perceived qualities of persons, competences, tools, methods, services, and artifacts involved in the learning process. Dimensions that define a learning space depend on the area of the learners’ interest, previous experiences and the intended learning outcomes. For example, a learning space for learning addition of natural numbers (positive integers) can have three dimensions – “number of summands” (having values 2 for two summands and 3 for more than two summands), “number of digits” (1 for one digit, 2 for two digits and 3 for more than two digits), and “method of calculation” (1 for using a calculator, 2 for written and 3 for mental calculation).

In a more formal framework, a learning space can be conceived of as a dynamic spatial ontology defined by ontological dimensions, in accordance with Kaipainen, Normak, Niglas, Kippar & Laanpere (2008). This is why we may call the learning space learning onto-space and refer to its dimensions as onto-dimensions. Accordingly, the coordinates of an element in learning space are called onto-coordinates. Learning space $A$ is described by an open-ended coordinate system $O = [x_1, x_2, ..., x_m]$. Each element of $A$ is represented by an $m$-tuple $A_i = (a_{i1}, a_{i2}, ..., a_{im})$ of numerical coordinates where $a_{ij}$ stands for the salience of $j$th quality with respect to the element. Semantically, meaning of $a_{ij}$ depends on the nature of dimension $x_j$, e.g., presence, proximity, probability, strength-of-relation.

Niche

We define a (learning) niche as a subspace of a learning onto-space defined by a subset of onto-dimensions and ranges (areas of values) within them, such that a community of learners and teachers with similar motives select in the learning process depending on a chosen goal. For example, if the goal is to learn summation of two natural numbers, the niche is determined by the value 2 of the second coordinate (the number of summands).

Formally, a niche $N$ in a learning space $A$ can be presented as a subset of $A$. A niche is both action- and meaning based, and can also be determined, for example, by certain aspects of identity of a learning community. Therefore, a niche can be a loose subset and it may lack a formal determination. The utilization of the concept of a niche is adopted from biology where it is extensively used to describe a region in an abstract space of environmental factors, in which a species has optimal living conditions for performing actions related to their life. Although Gibson (1979) previously considered niches in ecological psychology, the term has been used in the context of ecological learning systems only recently (Pata, 2009a, b).

Perspective

A learning space has, by default, too many dimensions to be made sense of directly. The learner can consciously perceive or pay attention to only a very limited number of dimensions at a time. A perspective is defined as a vector of weights of the learning space, one for each dimension with the extreme values of 0 standing for ignorance of the dimension and 1 for its full taking into account. Fixing a proper perspective the observer applies only a subspace of the whole learning space. Choosing certain weights for dimensions, a perspective produces a prioritization. When taking certain goal-directed actions, this allows focusing on one or on few dimensions of the learning space only. If, for example, the method of calculation is not important, the corresponding perspective is defined by vector $(1, 1, 0)$. 

264
Formally, a perspective $W$ of an $m$-dimensional learning space $A$ is a vector $W = (w_1, w_2, \ldots, w_m)$, where $0 \leq w_j \leq 1$ for each $j \in \{1, 2, \ldots, m\}$.

When shared by more than one individual, a perspective becomes community-defining, facilitates some community actions more than the others, and contributes to the determination of an abstract community-specific learning niche.

**Position**

An element of a learning space is called position; a position is represented by a vector of its onto-coordinates. For each learner, we attribute a specific position in the learning space (or in the niche) that determines to what extent the data items corresponding to the coordinates are currently relevant to the learner. The position attributed to a learner is referred to as the position of a learner. In the case of a “learner” consisting of a group of individuals, if a dimension describes the availability of a certain affordance (e.g., sharing artifacts, for the definition of affordance see below), the coordinate value 1 corresponds to the case in which this affordance is perceived by all members of the group and the value 0 to the one in which nobody from the community perceives it. A learner who is able to find a sum of only two natural numbers both having only one digit and only by using a calculator, has the position (2,1,1).

Although the learners are not considered as elements of a learning space we may project learners into it by identifying every learner with their position in the learning space: the fact that the position of learner $\alpha$ equals $P$ can be written as $L(\alpha) = P$; $L(\alpha^k) = P_k$ means that after performing $kth$ step in their learning path, the learner $\alpha$ reached position $P_k$.

**Learning objective and outcome**

A learning objective is defined as an intended learning outcome; it can be, for example, a material object, plan, idea or a competence. Learning objectives motivate learners to perform learning actions. The learning outcome is the result of these actions – increase of competences, a knowledge artifact or, for example, a completed project.

An ecological approach considers that an individual learner is influenced by the niche, which is determined by a learning community. The set of possible positions $H$ achieved by realization of a learning objective $O$ may consist of only one position or can cover a whole subspace of the learning space $A$. In our example of addition, the element (2, 2, 2) and the subspace (2, 2, x) with $x$ ranging over all possible values, would present examples of learning objectives. Therefore, we can define the outcome function $F$, setting $F(O) = H$ where $H \subseteq A$.

Note that the terminology used in the literature depends on what level the learning is considered. For example, according to activity theory (Kuutti, 1995), three levels of abstraction can be distinguished: activities, actions and operations. Activities are governed mainly by motives that are described in rather general terms, while on an action and operation level the objectives are usually expressed in terms of concrete goals.

**Step**

A step is defined as any event causing a change of a position of a learner in the learning space. Each step leads from the current position $P_{i-1}$ to the next one $P_i$. A step $S_i$ can be considered as an operator having the position $P_{i-1}$ as its argument and the next one $P_i$ as the result: $S_i(P_{i-1}) = P_i$. Here we may consider the activity theory notion of mediators of action including learning software, learning content, other people involved in the learning situation and rules and regulations in this community (Engeström, 1987). The mediators used while performing a step can be indicated by attributes of the corresponding step operator. In the example of addition, a teacher, a textbook, an internet site, a friend or a software based recommender can serve as mediators during a step that leads from the position (2, 2, 1) to the position (2, 2, 2).

**Path**

A path (or learning path) is defined as a chain of subsequent steps. Having fixed an initial position $P_0$ and a learning path $(S_1, S_2, \ldots, S_n)$ the learner achieves the final position $P_n = S_n(P_0)$ through intermediate positions.
\[ P_1 = S_t(P_0), P_2 = S_t(P_1), \ldots, P_{n-1} = S_t(P_{n-2}). \]

In the example of addition, steps leading consecutively through the positions \((2, 1, 1), (2, 1, 3), (2, 2, 1), (2, 2, 2)\) form a path; for example, a calculator, mother, or teacher respectively, can serve as mediators used while performing the steps.

For deciding what path should be chosen, different criteria can be used, for example:

- time for completing the path (total time for performing the steps)
- length of a path (number of steps)
- costs for actualizing necessary mediators (for example, buying a textbook)
- a weighted criterion.

A formal model for describing learning paths is presented in (Janssen et al, 2008).

**Learning pattern**

A *learning pattern* is a formal model for describing the general structure of a learning process by means of steps and positions in which learning objectives and various human and material resources in the learning space are interrelated and influencing each other for achieving a certain learning outcome. The approach builds on the concept of *pattern language* (Alexander, Ishikawa & Silverstein, 1977). It assumes that a pattern is a proven solution to a problem in a context, and that it is based on problems of fixed format and their solutions. Representation of a learning pattern depends on the modelling or descriptive tool used.

It is important to note that *learning pattern* and *learning path* are fundamentally different concepts. A learning pattern \(Q\) is formed due to the continuous generalization of a set \(Q\) of many learning paths having similar objectives and can be achieved through factorization of the set of learning paths \(Q\) by a certain similarity relation \(\rho\), formally, \(Q = Q/\rho\). Therefore, a learning pattern allows a variation of different paths that the community members have frequently used for achieving certain learning objectives, whereas a learning path describes meaningful ways for a concrete learner to achieve the outcome in given conditions. The actual learning process forms a learning path; the latter can be considered as a realization of some learning pattern. Therefore, learning patterns serve a learner as general guides for deciding on next steps.

Let Figure 1 represent the graph of all steps that appeared in a hypothetical learning addition of natural numbers. The positions of a learner that are considered similar are surrounded by dashed boxes:

![Diagram](image)

*Figure 1. The graph of steps in addition of natural numbers*

Identifying similar positions (that is, considering the set of mutually similar positions as one position), we can identify the learning pattern. Therefore, the learning pattern of addition of natural numbers can be represented by the following directed graph (using corresponding notations for vertices of the graph):
Both the graph of steps and the learning pattern can be applied for different niches as well. For example, if only a calculator and mental calculation are possible (meaning that the third coordinate can not have the value 2), we obtain a corresponding learning pattern by leaving out vertices P₅ and P₇ together with the adjacent edges.

Should the learner, for example, neglect the method of calculation – this is determined by the perspective (1,1,0) meaning that the two positions will be identified exactly if their first two coordinate coincide – the positions in subsets \{P₀,P₁\}, \{P₁,P₅\} and \{P₂,P₆,P₇,P₈\} would be considered as equivalent and the learning pattern would reduce to

\[
\begin{align*}
P₀ & \quad P₁ \\
| & | \\
P₂ & \quad \quad \\
\end{align*}
\]

Figure 3. The learning pattern for the addition of natural numbers in the case of neglecting the method of calculation

Here P₀ denotes the position "adding two one-digit natural numbers," P₁ denotes "adding two natural numbers (of arbitrary length)" and P₃ denotes "adding more than two natural numbers (of arbitrary length)."

**Direction of a step**

*Direction of a step* is defined as a relative position of the next position with respect to the current position without considering the distance between the positions. Formally, if a step leads from position \(P_{i-1} = (a_{i-1,1}, a_{i-1,2}, \ldots, a_{i-1,m})\) to position \(P_i = (a_{1}, a_{2}, \ldots, a_{m})\) then its direction equals to the vector \(P_{i-1}P_i = (a_{1} - a_{i-1,1}, a_{2} - a_{i-1,2}, \ldots, a_{m} - a_{i-1,m})\). Therefore, an arbitrary set of features causing a change of the position determines a certain direction. In learning, choosing the direction of a step is represented by perception and actualization of a set of dimensions of the learning space related to achieving a certain goal and corresponding mediators of learning. For each step, there normally exist a number of possible directions, which lead to different outcomes (to different next positions of a learner in the learning space) depending on the attributes of the step. Here again, the direction of a step is determined only if the position of which the step is applied to is fixed. Figure 4 below, illustrates the relationship between the concepts of position, step, path, learning pattern and direction of a step.

The step leading from position (3,1,3) to position (3,2,2) has (0,1,-1) as its direction. This means that the length of the summands has increased and the complexity of calculation has decreased by one level while performing the step.

**Step gradient**

*Gradient (or step gradient)* is defined as the most preferred direction of the learner toward the next position in a learning path, as seen from the currently selected perspective. Following the gradient offers the learner an opportunity to make the biggest advancement toward learning objective.

Gradients are dependent not only on varying perspective but also on other actors taking actions in a particular learning space. We can put it even more strongly: the gradient is not only a function of learning objectives but also – and possibly to a greater extent – of the current position of the learner and learning opportunities offered by the learning space.
For example, after having learned how to use a calculator for adding two one-digit numbers (that is, having the position \((2,1,1)\)), the most preferred direction would be \((2,1,3) - (2,1,1) = (0,0,2)\) if a calculator is not available in later stages of learning addition. On the other hand, should a calculator be the main tool for checking the accuracy of the calculation, the most preferred direction would be either \((2,2,1) - (2,1,1) = (0,1,0)\) or \((3,1,1) - (2,1,1) = (1,0,0)\), depending on the learning objective.

**Affordance**

We define an *affordance* as a perceived action-promoting property or relation between particular aspects of the situation and the subject who plans or undertakes actions in a certain environment.

More specifically, we see *learning affordances* as something that learners with certain learning goals belonging to a community culture, perceive when they interact with the components of the learning situation (e.g., human and material resources, tools) in certain particular learning environments. The accumulated set of affordances perceived by the majority of the community members can characterize the learning affordances of this community culture (Pata, 2009a, 2009b). Individual members of the community can use such a set of community’s learning affordances as a guidance to interact with their learning environment in order to achieve their learning objectives or pursue their learning motives in accordance with those of the community. Affordances provide the cognitive basis for our approach because they convey the idea of cognition that enables an individualized approach in goal-directed action.

**Moving towards the learning objective**

After we have defined the basic notions we are now able to describe the ecological approach to learning that enables a learner to be influenced by the dynamic learning space and simultaneously contribute to the formation of this space. The key question for a (self-directed) learner is to find a path that leads them from the current position in a learning space to the target position or to a region determined by the learning objective. Depending on the learner’s preferences and imposed restrictions it should be possible to choose between different learning paths.

As an example of a learning path to follow, the learner may:

1. Determine the learning objective.
2. Decide on the sub-goals and their order.
3. Select the first sub-goal and determine the niche within the learning space suitable for achieving the sub-goal.
4. Decide on a step toward the selected sub-goal in the gradient direction (for example, by selecting an appropriate mediator of the learning event).
5. Perform the step.
6. Assess whether the sub-goal is reached; if not, return to 4 (decide on a new step).
7. Assess whether the learning objective has been reached; if not select the next sub-goal, determine the niche and return to 4.

As mentioned before, the most preferred direction of a step (the step gradient) is influenced by both the initial and target positions as well as learning opportunities. There are several possibilities for determining the direction of the step, for example, following suggestions of a tutor or analysing an appropriate existing learning pattern.

Figure 4 illustrates the navigation of a learner in the learning space that has accumulated a certain learning pattern. Each arrow represents a step – the thick arrows represent a learning pattern, thin continuous arrows represent the steps along the gradients and dashed arrows represent the other possible directions of steps for a learner. A possible strategy of using learning patterns for determining a learning path can be the following: initially (during the first steps) the learner aims to get close to a node of the learning pattern, while subsequent steps will be made according to the learning pattern.

In Figure 4, the first step brings the learner from initial position \(L(\alpha^0)\) close to a node of the learning pattern; the second and third steps (from position \(L(\alpha^1)\) to \(L(\alpha^2)\) and to \(L(\alpha^3)\)) already follow the learning pattern. Although direction \(d_2\) would lead into the outcome area \(F(O)\) as well, the learning pattern does not suggest it. Whether to
choose another possible direction $d_i$ suggested by the learning pattern or not depends on additional circumstances (see the case study below).

![Figure 4. Using the learning pattern to find a learning path](image)

**Case study**

As an example we take the empirical data from Pata (2009a). A group of 53 master students at Tallinn University participated in the course “Self-directed learning with social media.” We demonstrate how the learning space and niches for individual and collaborative activities were determined with the purpose of guiding a learner.

We first determine the learning objectives. Then we determine the set of dimensions of a learning space and describe the mediators (possible learning tools) by means of these dimensions. Then the standard multidimensional scaling algorithm, e.g., Kruskal & Wish (1978) is applied in order to calculate a cluster map on which the similarity of mediators is expressed by means of their proximity of their images on the map. Using this map, one can identify the mediator that is most suitable to apply for performing the step.

**Determining the learning objectives**

The main learning aim of the course was to develop self-directed learning competences through individual and collaborative assignments with social software tools. As an individual assignment each learner composed a personal learning environment (PLE) for supporting their learning activities, and described its affordances. Learners were permitted to use various social media tools of their own choice; blog was the only mandatory tool to centrally monitor the progress of learners. As a collaborative assignment, the learners connected their PLEs with some associated social software tools, composing a distributed learning environment for conducting collaborative assignments. Again, they had to perform some activities in this environment and describe the affordances of the learning environment. These two learning objectives – to compose a PLE and to connect them – were determined by the teacher; however, the students had the freedom to choose the tools, methods of assembling the tools together, and situating the assignments into their particular work or learning context.

**Determining the dimensions of the learning space and identifying the mediators of learning**

Each learner described a set of affordances that they perceived when doing individual or collaborative assignments with certain tools. After the course these affordances were collected into a dataset that was used as an example of an authentic learning space for self-directed learners using social software. The affordances were grouped into categories to minimize the system complexity. Initially the following 19 affordance categories were formed: assembling, managing, creating, reading, presenting, changing, collaborating, sharing, exchanging, searching, filtering/mashing, collecting, storing, tagging, reflecting, monitoring, supporting, asking/feedback, and evaluating. Each category of affordances was taken to represent an onto-dimension and collectively they would constitute the learning onto-space.
Altogether 12 types of software were considered as tools in PLEs and collaborative learning environments: blog, wiki, chat, social bookmarking, aggregator, email, search engine, co-writing, forum, co-drawing, Flickr, and YouTube.

Each tool was assigned a 19-dimensional vector with coordinates within range 0…1 estimating to what degree the particular tool offered corresponding affordances in case of individual assignments compared with other available tools. For determining the coordinates, the frequency of how often each affordance/dimension was perceived while using a particular tool type in the sample group was calculated: the total number of tools used per dimension constituted 100%, from which the actual frequency of perceiving the usefulness of the tool was calculated. For example, the affordance assembling was mentioned by 49 students, while only 20 students perceived this as a blog affordance – therefore the assembling coordinate for blog was given a value equal to 20/49 = 0.41.

Setting sub-goals

As already mentioned, the following sub-goals were set: 1) composing a PLE and using it for an individual assignment and 2) connecting the PLE with the PLEs of other learners for managing a collaborative assignment. According to these goals, two niches were considered, one for performing the individual task (composing a PLE) and one for performing the collaborative task (connecting the PLEs).

For illustration purposes we will now describe a simplified learning path for one hypothetical learner consisting of two steps where the goals will consecutively be reached by one step only.

For the representation of individual and collaborative learning niches, as well as finding the most appropriate learning tools, we used the Onto-space Explorer tool (OSE tool; Kaipainen et al., 2008). This tool applies multidimensional scaling to compute a cluster map, allowing users to control the weight given to each coordinate by means of a slider (Figures 5 and 6 below).

Figure 5. The results of selecting the perspective \{managing = 1, creating = 1, all other weights = 0\}
Determining the first step

First, the learner should identify the affordance requirements for achieving the first goal. Let us assume the affordances managing and creating were chosen. Thus, a perspective \{managing = 1; creating = 1, and for all other dimensions the weight = 0\} was fixed. Fixing the perspective the learner prioritizes the affordances managing and creating for managing personal tools and using the PLE for an individual creative assignment (composing the PLE and analyzing it).

In our case, the learner determines a perspective by shifting the sliders for the dimensions managing and creating to 1 (see Figure 5). On the graphical map generated by the OSE tool, suitability of the learning tool is measured by the distance of its image from the upper right corner of the map; we find that blog – indicated by the block arrow – is the most suitable one to use for performing the first step \(S_1\) (note that forum and wiki are so equally unimportant that their images almost coincide on the map).

Now two possible cases may hypothetically appear:
Case A: If a learner already is a blog user, no correction is needed in their first learning step for doing individual assignments.
Case B: If a learner had previously used, for example, wiki for creative tasks, they might appear distant from the community’s activity niche for doing individual assignments.

In both cases, the learner must choose the most preferred tool, that is, in terms of the model to find the gradient of the step. In Case A, the most preferred tool is most likely the blog. In Case B they can keep using wiki only if it will be possible to get into the community niche during the second step (that is, if a tool exists that allows them to connect their PLE with the PLEs of other learners for managing a collaborative assignment). Otherwise they should acquire the skills for using blog as well and start using it.

![Figure 6. The results of selecting the perspective \{assembling = 1, managing = 1, all other weights = 0\}](image)

Determining the second step

In performing the second step \(S_2\) for managing a collaborative activity the learners selected assembling and managing as the most important expected affordances using the collaborative assignment niche. During this step the
group members’ PLEs should be assembled into a distributed learning environment, and afterwards managed. Consequently, the perspective \{assembling=1, managing=1, all other weights=0\} was chosen. The results computed by the OSE tool are presented in Figure 6. The indication is that blog and aggregator are suggested as the most preferred tools (note also that only chat and wiki are of some importance for completing collaborative assignments; all other tools are equally unused).

The learning pattern

The composition of the learning pattern that is based on the empirical data in the example is presented in Figure 7 below. This pattern suggests three possible learning paths for moving from the initial position (situated in a niche $N_1$ for performing an individual task) to a second position (situated in a niche $N_2$ for performing a collaborative task).

The selection of a path depends on learner’s preferences and imposed restrictions. A blog user can continue using blog, adding RSS feeds from other blogs, or they can additionally start using an aggregator for pulling RSS feeds from other learners’ blogs for monitoring their reflections (Path 1). A wiki user can either begin by acquiring skills for using blog, build a PLE using blog and then start using an aggregator (Path 2), or keep using wiki and then start using an aggregator (Path 3). However, Path 3 may be less effective if all the other learners are using aggregated blogs for monitoring purposes.

![Figure 7. Learning paths for individual and collaborative assignments](image)

Discussion

Previous approaches to learning design have typically focused on the developers’ or instructors’ view to learning systems and did not consider the possibility of learners to change dynamically its actions depending on perception of the system and behaviour of the learning community. Although that some authors have proposed affordance frameworks and classifications for solving particular tasks, for example, the 24 features identified by Palmer, Sire, Bogdanov, Gillet and Wild (2009) for mapping the functionality of web PLEs, no holistic computational model of describing learning affordances and dynamics has been suggested until today.

It is our experience that, where learners are provided with the freedom to choose their own tools and are encouraged towards self-directed learning, the teaching process requires appropriate decision support instruments (Fiedler & Pata, 2009). Such instruments (for example, a learning pattern or the OSE tool) enable the dynamic formation of learning spaces that are defined by the learners themselves, such that they can be used for personal or group navigation within the learning space.

Our approach is based to a great extent on the concept of the niche as a subspace of a learning space that can be dynamically determined. The fact that a niche can be action or meaning-based and may lack a formal determination can sometimes be truly challenging. Moreover, behaviour of a learner in a niche depends on a chosen perspective. Consequently, the intermediate positions of different learners for achieving the same learning objective can be different. On the other hand, this offers an opportunity to plan the learning activities and learning steps in a situation where the abilities and conditions for individual learners are different.

The proposed approach offers each individual learner two major instruments for planning and conducting their learning: (1) a possibility to contribute to the learning space as it was demonstrated in the case study (dimensions of
the learning space – affordances – were determined by participating students) and (2) a possibility to follow personal priorities (by selecting the appropriate perspective) and to adapt learning to different types of activities (by selecting an appropriate niche) as it was demonstrated by the example of addition with natural numbers.

Furthermore, we believe that application of OSE tool type of navigational decision tools that collect information about learning activities from social software systems would allow novel learning design approaches (such as for example swarming phenomena that are based on self-organization and self-regulation of PLE users).

Notice that the definition of learning space offers a possibility for considering different types of learning spaces: the coordinates of a learning space can describe the learning object (the example of summation of naturals), the qualities of learning tools (the case study above), or the qualities of a learner. Therefore, for example a component of a PLE can be represented by a single coordinate (a calculator in summation), by a set of coordinates (the case study), or it can be considered as a mediator of action and represented as an attribute of a step. Consequently the approach can be applied in different contexts and for solving different types of problems.

**Conclusion**

This paper presents a formal model of learning, based on a spatial metaphor. As the fundamental element of this model, an ontological space of affordances is proposed. It assumes a collaborative annotation practice, similar to many applications in social media. It constitutes the coordinate system for describing emergent community learning space, and at the same time, the framework for creating and applying personal learning paths and finding learning patterns. Beyond merely hypothetical modelling, we believe this model has potential for contributing to the methodology of learning design. This approach allows planning and analyzing learning designs from different perspectives as a result of different weightings associated with each dimension of the learning space. The software applications that implement the ideas of this model also allow for new designs for learning in PLEs. As this approach is not bound to any particular level of abstraction, in principle it can be applied equally in the analysis of general educational processes, as well as for guiding concrete learning activities.

**Acknowledgements**

The authors thank Viive-Riina Ruus for interest in the study and stimulating discussions as well as the anonymous referees for the many suggestions that contributed to improvements in the exposition of this paper. This study was partly funded by Estonian Science Foundation grant 7663 and by Estonian Ministry of Education and Research grant 0130159s08.

**References**


KnowledgePuzzle: A Browsing Tool to Adapt the Web Navigation Process to the Learner’s Mental Model

Iyad AlAgha

Department of Engineering and Computing Sciences, University of Durham, UK // i.m.q.alagha@durham.ac.uk

(Submitted November 18, 2010; Revised March 2, 2011; Accepted May 16, 2011)

ABSTRACT

This article presents KnowledgePuzzle, a browsing tool for knowledge construction from the web. It aims to adapt the structure of web content to the learner’s information needs regardless of how the web content is originally delivered. Learners are provided with a meta-cognitive space (e.g., a concept mapping tool) that enables them to plan navigation paths and visualize the semantic processing of knowledge in their minds. Once the learner’s viewpoint becomes visually represented, it will be transformed to a layer of informative hyperlinks and annotations over previously visited pages. The attached layer causes the web content to be explicitly structured to accommodate the learner’s interests by interlinking and annotating chunks of information that make up the learner’s knowledge. Finally, a hypertext version of the whole knowledge is generated to enable fast and easy reviewing. A discussion about the affordances of the tool and how it can be used to achieve a scaffolded approach for learning is presented. The evaluation of the tool in a real educational setting revealed its support for navigation planning and knowledge recall as compared to classical navigation techniques.

Keywords

Navigation planning, Web annotation, Concept map, Mental model, Learning technology

Introduction

The web is an open information system in which various information resources can be interlinked together in diverse ways to form hypertext or hypermedia environments. The ability to structure and integrate various learning materials makes the web a rich educational medium. People usually navigate the web in a self-directed way. They direct their navigation routes differently in accordance with their individual preferences. They may select and examine from a large pool of information only those pieces necessary to meet their objectives. Thus, the structure of knowledge gained from the web is likely to be different from one user to another due to the different navigation paths they follow to gather information. This structure may also differ from the structure of web content since information may be obtained from diverse documents which are not directly linked. Therefore, hypertext documents, due to their static nature, cannot cope with individual differences among self-directed users. They cannot be interlinked or restructured to match the knowledge representation in the mind of every single user, unless the user is himself/herself the content author. This gap between the user’s information needs and the structure of information on the web can place navigation and cognitive overload on users due to the difficulty of monitoring their progress and recalling information components while navigating the web (Conklin, 1987).

To illustrate the above problem, imagine the scenario of a learner navigating the web to learn about a particular topic. The learner accesses various web pages, each of which belongs to a different site and explains a different aspect of the desired topic. While navigating, the learner constructs knowledge by rethinking about the navigated content and making semantic links between pieces of information scattered on various pages. The fact that these pages are not organized on the web in the way that the learner demands causes him/her to do additional cognitive and navigational effort to retrieve information of interest and retain the structure of knowledge in mind. This effort can cause the learner to fail in knowledge construction if large amounts of information need to be retained.

It is obvious from the above discussion that there is an emerging need for techniques that enable users to structure and interlink information on the web in proportion to the progress of their own navigation process, rather than being typical viewers of information. The vision of this work is to propose a technique which aims to adapt the navigation path on web to reflect the mental representation of knowledge by utilizing an information visualization technique. Learners are provided with a meta-cognitive space (e.g., a concept mapping tool) that enables them to visualize how they like information resources to be structured. Subsequently, the constructed visualization is converted to a hypertext layer over the visited pages, causing the information structure on the web to more accurately represent the
knowledge structure in the learner’s mind. This can considerably enhance accessibility and thus reduce cognitive overload. A thorough discussion about the proposed technique and how it can be used to achieve a scaffolded approach for learning is presented. The evaluation of the technique using both quantitative and qualitative measures within a formal learning setting is also explained and results are discussed.

Self-directed learning from the web

The term “self-directed learning” originated in the field of adult education (Roberson, 2005). It has been defined as “a process in which individuals take the initiative, with or without the help of others, to diagnose their learning needs, formulate learning goals, identify resources for learning, select and implement learning strategies, and evaluate learning outcomes” (Knowles, 1975). With the growing trend toward web-based learning, the concept of self-directed learning has received increasing attention. Hanna et al. (2000) stressed that it is a key factor in successful web-based learning. Cennamo et al. (2002) found that success in web-based courses often depends on the learners’ abilities to successfully direct their own learning efforts and to decide on suitable navigation paths. Learners in online environments who are skilled in self-direction become more responsible for their learning and more self-motivated (Chang, 2005). However, these positive effects of self-regulation have to be balanced against the problems of learning from the web, and in particular the problems of disorientation and cognitive overload.

A prerequisite for designing a web-based tool to promote self-directed learning is to understand how learners construct knowledge while navigating the web. Using the experience from previous research (Kashihara & Hasegawa, 2004; Mitsuhara et al., 2008; Tergan & Keller, 2005), this section illustrates how the human brain retrieves and processes information while browsing the web. This will be a prerequisite for visualizing the navigation path based on the learner’s mental model at a later stage.

Figure 1 depicts an example of how knowledge is structured in both the web and the learner’s mind. Learners often start navigating the web with a particular goal in mind. For example, they may search for a particular definition or a description for a specific topic. They navigate from one web page to another until they find the information that answers the initial learning goal. In Figure 1, information that meets the learner’s interests has been located in three different pages that are not linked via direct hyperlinks (components A, B and C). The learner builds knowledge by creating semantic links between these disparate pieces of information. Each link denotes the goal of navigating from one component to another. For example, the learner may think that component B illustrates the information learned from component A, and that component C supplements the description given in component B. However, these links only exist in the learner’s mind, and the original pages are not explicitly linked in a way that conforms to the learner’s mental model.

It should be mentioned here that the learning goal arising from browsing a web page is not always answered in the immediately following page. In addition, it is likely that further learning goals will arise while looking for the answer of a retained goal. It is also possible that the learning goal may have one starting page with multiple target pages. For example, the answer to a particular question could be found in the combination of information assets collected from various resources. All these cases require the learner to make extra cognitive effort to manage the navigation process, retain learning goals and recall information of interest from previously navigated resources. Our hypothesis is that this cognitive effort can be mitigated if the links that the learner mentally creates while navigating the web can be physically converted into informative hyperlinks and annotations on the web. The KnowledgePuzzle tool was built to achieve this goal. The tool and its functionalities are demonstrated in the following section.
The KnowledgePuzzle tool

In order to help learners visualize the navigational learning process, the KnowledgePuzzle tool shown in Figure 2 was developed. The tool’s main window is split into two spaces: a browsing space (on the left) and a planning space (on the right). The learner uses the planning space to visualize the cognitive model of knowledge which was constructed while browsing the web. Accordingly, resources on the browsing space will be interlinked and annotated to reflect the constructed visualization.

![Figure 2. The knowledgePuzzle tool](image)

Knowledge visualisation

The planning space enables learners to assemble a sequence of web pages that fulfill the desired learning goals. It emulates the navigational learning process explained in the previous section: when the learner decides to set up a learning goal in reading a page, he/she can add a graph node denoting the page, or a paragraph within the page, where the goal arises. Similarly, when a retained learning goal is fulfilled, another graph node can be added to denote the target page, or a portion of it, where the learning goal is achieved. Adding nodes is simply done by highlighting parts of interests from the browsing space, dragging and then dropping on the planning space. Nodes can be optionally named with the main concept learned from the source pages. To visualize the relationships between pages, links can be drawn between nodes and annotated with terms indicating the goal of navigating from a page to another. Each graph node is represented as a thumbnail of the source page. In addition, each graph node is hyperlinked, and clicking on it causes the source page to open in the browsing space and the part of interest to be highlighted. This direct linking between graph nodes and information assets in hyperspace enables rapid access to information and helps learners to consolidate the correct sequence of pages at any time.

Learners can manipulate the constructed graph during the navigation process whereas each manipulation is executed by means of mouse clicking or dragging. There are four basic manipulations: adding, editing, deleting and altering the navigation goal links between nodes. Zooming techniques are also supported to facilitate manipulation of large constructed graphs: the user can zoom in to focus on specific parts or zoom out to have a global view. Notes about the content of any page can be inputted through the entry form that will pop up when clicking on each graph node. Later on, these notes will be attached to the related content within the web page so that the learner can review the page along with his/her notes.

Figure 3 shows a sample graph constructed for the goal of learning the photosynthesis process in plants. It depicts the nodes representing the web pages being visited and the primary navigation processes that have been executed. Each graph node is named by the learner with the main topic learned from the source page. For example, the learner
visited the pages named “Chlorophyll,” “Carbon Dioxide (CO2)” and “Oxygen” to learn about the requirements and products of the photosynthesis process which was explained in a previously visited page. After learning about the “Carbon Dioxide” from a particular page, the learner obtained supplementary information from a page named “Biological role of CO2.” Links between nodes indicate the goal of navigating from one page to another. For example, the link labeled “requires” denotes that the learner visited the destination page to learn about a requirement of the concept explained in the previous page. When the learner creates a link between two nodes, both the relationship and its inverse should be inputted. For example, if the link from “Photosynthesis” to “Chlorophyll” is named as “requires”, the inverse link should be named as “is required by” or something similar. While, for simplicity, inverse links are not shown, they will be used when the graph is converted to a hypertext layer over web pages in the next stage.

![Figure 3. A sample knowledge graph](image1)

![Figure 4. The conversion of the knowledge graph to hypertext components: (A) an excerpt of the knowledge graph, (B) The constructed hypertext menus](image2)

**Hypertext layering**

After visualizing the learner’s navigation path, the tool automatically transforms the graph to a layer of hyperlinks and annotations which will be laid over the visited pages. The layering process is done as follows: The links connecting any source node with target ones are transformed to a list of hyperlinks. The constructed list is embedded inside the source page, aiming to link information of interest within the source page with related information in target
pages. Similarly, lists of inverse hyperlinks are created and embedded inside target pages to denote inverse relationships that link back pieces of information in target pages with source pages.

Figure 4 shows an example of the conversion process. Figure 4.A depicts two graph nodes extracted from Figure 3 with the associated relationships. Figure 4.B shows the corresponding lists whereas each list includes hyperlinks denoting the relationships associated with each node. Since the original pages on the web cannot be altered, the new structures will be laid over the copies of web pages after being loaded on the browser.

Figure 5 shows how the web pages, represented by the graph nodes shown in Figure 4, look like after the new layers of hyperlinks are attached. The part of the page that the learner is interested in is formatted in a different color and a new hyperlink labeled “Click here” is added to its end. Clicking on the link causes the list of hyperlinks to be displayed. These hyperlinks link to target pages and expose how they are related from the learner’s perspective. Therefore, they enable seamless transition between resources of information that make up the learner’s knowledge.

The naming of the attached hyperlinks enables learners to recall what was the goal achieved by navigating from a page to another, thus providing support for learners to reconstruct their cognitive models. In addition, both lists include a link labeled “My notes” on which clicking will open a small window showing the notes added by the learner about this part of information. This enables learners to review their comments and notes alongside the related web content. On visiting any page, the tool automatically builds the lists of hyperlinks, according to the structure defined on the planning space, and then attaches them without imposing any significant delay. Thus, it will appear to the learner as if the original page is personalized to accommodate his/her own needs. One should note that the original web pages do not include these changes and thus they are not initially structured to match the learner’s viewpoint. The tool was developed using a Java-based web browser component (ICEBrowser, 2011) which...
implements portions of the W3C Document Object Model (DOM) Level 1 specification. This enables to programmatically access and manipulated the web pages loaded into the browser for the purpose of highlighting the parts of interest and attaching the list of generated hyperlinks.

After structuring the web pages to match their information needs, learners can hide the planning space and focus on the navigation process through the browsing space, which becomes guided by the newly attached hyperlinks and annotations. They need to open the planning space only if they want to define new sequences of pages or alter existing ones. The attached hypertext layer is constantly synchronized with the graph on the planning space: changes made on the graph by the learner will be instantly applied on the browsing space.

Generation of hypertext knowledge

In addition to building new associations of web resources, the tool produces an independent hypertext representing the whole constructed knowledge. The generated hypertext is a single page that contains all referenced pieces of information interlinked as in the knowledge graph. Figure 6 shows an excerpt of a sample page generated for the graph in Figure 3. Referenced pieces of information are extracted from source pages and embedded in the generated hypertext (see elements A1, A2 and A3 in Figure 6). The page is hierarchically structured so that the target pieces of information become subsections of the source ones. The user-define relationships between nodes are also revealed so that the learner can recall how he/she semantically processed and linked information resources while learning (see elements B1 and B2 in Figure 6). User’s notes and comments are also attached to related components (see element C1 in Figure 6). The generated hypertext enables rapid reviewing of knowledge by gathering user’s notes as well as all pieces of information from visited pages, and organizing them in a single page based on the user-defined relationships.

The constructed knowledge can be exported to a file so that it can be imported into the tool and reused later. The graph is implicitly converted to an XML format in order to be machine-processable. Importing such a file will not only rebuild the hypertext format, but will also reconstruct the complete graph on the planning space. This supports persistence of the learning session for resumption by the learner at a later time.

Discussion: the use of the KnowledgePuzzle tool for self-directed learning

After the functionalities of the KnowledgePuzzle tool have been demonstrated, the following section briefly discusses how the tool can be used to achieve a scaffolded approach for self-directed learning from the web. Figure 7 depicts an overview of the learning stages and how the KnowledgePuzzle tool is utilized to facilitate learning at each stage.
Stage 1: Navigation planning and monitoring

This stage involves learners when they start navigating the web with the aim of gathering information. At this stage, learners need support to monitor and control the learning process. While browsing, they use the planning space to plan which page, or part of a page, to visit and the sequence of pages to be visited so that the learning goal is achieved. Subsequently, information components on the web can be previewed according to the planned path. From an educational perspective, helping learners build the optimum path that fulfils their goals can release them from unnecessary browsing activities. The separation between the navigation path planning and the web exploration processes at this stage is crucial as it allows learners to become aware of monitoring their navigation process. Learners use the planning space to practice meta-cognitive skills such as planning, monitoring and revising the information gained from the web. If the learners become able to externalize what they have in mind and visualize it in an understandable format, they will be released from the cognitive effort required to retain the acquired knowledge in mind.

Stage 2: Transformation of the navigation map to hypertext

This stage involves learners when they start rethinking and reviewing the knowledge gained during the exploration process. Learners at this stage need to revisit pages that they found useful during the first stage. They also need to recall why they visited these pages, which page contents were important, and how different pages were semantically related. In order to assist learners to achieve that, the links connecting the sequence of pages on the planning space are converted to real hyperlinks and annotations inside the pages. From an educational perspective, this conversion can help learners orient themselves in hyperspace in such a way that facilitates rethinking the knowledge without making excessive cognitive effort. In addition, by moving from the planning space to hyperspace, learners are no longer dependent on any extra tools that may cause distraction. They can focus their attention on the value of information being browsed, which becomes explicitly interlinked to accommodate their interests.

Stage 3: Knowledge generation in hypertext format

This stage enables learners to automatically transform the constructed knowledge into a hypertext format. Such transformation aims to promote self-reflection by bringing the whole structure and information into the learners’ view, and thus enabling them to review their knowledge rapidly without the need to revisit the source pages.
It should be noticed that the above stages complement rather than substitute each other. Learners can move from one stage to another in accordance with the progress of the learning process. For example, they move from stage one to stage two when they think that they have finished the exploration and planning processes and now they need to start revision and reflection. The transition to the third stage provides an opportunity to put the whole knowledge in a single space of hypertext to facilitate self-review. However, if learners need to revise the relationships between knowledge components in the two upper stages, they can go back to the planning space in stage 1 and alter these relationships between the graph nodes.

Evaluation

An explanatory study was performed to evaluate the KnowledgePuzzle approach with the following objectives in mind:

- Ascertain if the KnowledgePuzzle tool facilitates navigation planning and page revisiting compared to traditional browsing techniques.
- Explore participants’ perceptions of the tool and their suggestions regarding its design and use.

Experimental settings

Participants were first-year undergraduate students at the department of Computer Science who were undertaking the “Computer Systems” module. A total number of 44 students (36 males and 8 females) participated in the study. Students were instructed to complete an assignment in which they need to study a collection of web pages with the aim of writing a short report that describes the security threat ‘computer viruses.’ The objective of the assignment was to gain knowledge of computer viruses as they relate to computer systems in general and to computer networks in particular. The students were told that the report should include the following sections:

- A general description of a virus, including how a virus spreads.
- A description of an instance of a virus.
- How the network is used and/or affected as a consequence of viruses.
- How to protect computer systems and users from viruses.

The task was deliberately designed to have multiple requirements as this would stimulate students to follow different navigation paths to complete the task. The learning material was a corpus of selected web pages related to the topic of computer viruses. Table 1 outlines about the corpus and indicates their complexity. Students were not asked to browse and read all pages but only those that they need to complete the task.

<table>
<thead>
<tr>
<th>Learning Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Pages</td>
</tr>
<tr>
<td>Number of Links per Page</td>
</tr>
</tbody>
</table>

Procedure

Two weeks prior to the experiment, the tool was demonstrated to the students and they were instructed to download and use it. The experiment was conducted in the lab where students were randomly divided into two groups. While one group completed the task using the KnowledgePuzzle tool, the other group used a traditional web browser to complete the task using the same corpus of resources. Groups were then switched and were given a different corpus, but with a similar complexity, to do the same task using either the KnowledgePuzzle tool or the traditional browser. Students were free to make decisions concerning the navigation paths they followed or the pages they needed to use. The duration of the task was 2 hours for each session.

During the experiment, navigation activities were being recorded in log files which were then collected and analyzed. Finally, a questionnaire was circulated. The questionnaire was divided into two parts: The first part
consisted of six multiple choice questions, shown in Table 2, each with a four-point scale (Strongly Agree, Agree, Disagree or Strongly Disagree). The second part required a written answer about any suggestions to improve the tool.

<table>
<thead>
<tr>
<th>Table 2: Questionnaire (Multiple-choice questions)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Questionnaire questions</strong></td>
</tr>
<tr>
<td>Q1 The tool helped me link separate pieces of information available on the web.</td>
</tr>
<tr>
<td>Q2 The tool helped me directly access information I needed inside the web pages.</td>
</tr>
<tr>
<td>Q3 The new type of links attached to the web pages by the tool helped me easily navigate through the web pages.</td>
</tr>
<tr>
<td>Q4 The tool effectively reduced the amount I need to remember.</td>
</tr>
<tr>
<td>Q5 Grouping and structuring information components in a single page helped me quickly review the knowledge gained from the web.</td>
</tr>
<tr>
<td>Q6 The tool was easy to use.</td>
</tr>
</tbody>
</table>

**Results and discussion**

In the next sections, we describe the results of the study. We begin by reporting analysis of navigation activity, and then we report results from the questionnaire.

**Analysis of navigation activity**

Navigation activities in both conditions were compared. Our assumption was that students opted to revisit pages of interest in order to rethink and organize information in preparation for the writing process. In case self-regulation skill is improved, navigational learning process is expected to become more convergent and goal-oriented. In other words, page revisiting for reflection on and reconstruction of knowledge would converge on restricted pages, and pages unrelated to learning goal achievement would not be frequently revisited.

Table 3 summarizes results in both conditions: using the KnowledgePuzzle tool, students visited an average of 21.5 pages. Of all visited pages, an average of 8 pages was actually used by each student to obtain information required for the task. These were the pages that the student referenced in the written report as well as on the planning space of the tool. We refer to these pages as the “instructive pages.” The average number of revisiting these pages accounted for 68.6% of the whole number of revisits. In the case of the traditional browser, students visited more pages (mean = 32.5) and executed fewer revisits (mean = 32.8). Only 39.3% of these revisits were related to the pages that the students referenced in the written report. T-test showed that the tool (KnowledgePuzzle and traditional web browser) had a significant main effect on the percentage of revisits to instructive pages (t(21) = 2.9, p < 0.05, two tailed paired T-test). This result showed that students revisited significantly more pages than necessary using the traditional browsing approach. In contrast, the high percentage of revisits to instructive pages in the case of KnowledgePuzzle tool indicated that it directed learners’ attention to the primary exploration processes by helping them to revisit pages only required for information retrieval. This made the navigation process more convergent and goal-oriented.

<table>
<thead>
<tr>
<th>Table 3. Analysis of navigation activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average number of different pages visited</strong></td>
</tr>
<tr>
<td>-----------------------------------------</td>
</tr>
<tr>
<td><strong>KnowledgePuzzle</strong></td>
</tr>
<tr>
<td><strong>Traditional web Browser</strong></td>
</tr>
</tbody>
</table>

Furthermore, we analyzed the KnowledgePuzzle utility to determine how learners used the tool to revisit web pages. Table 4 outlines the average number of revisits executed using each of the navigation aids offered by the tool. It shows that learners used three different ways to revisit pages: 1) by following the new hyperlinks attached over web
pages, 2) by clicking the hyperlinked nodes in the constructed graph or 3) by using the “back” and “forward” buttons. Results showed that the newly attached hyperlinks were the preferable way for learners to make the navigation path, followed by backtracking and graph nodes. This result supported our hypothesis that restructuring web resources to match user information needs and making this process smoothly integrated with reading facilitates revisiting of pages compared to classical revisiting techniques (e.g., backtracking).

<table>
<thead>
<tr>
<th>Hypertext Layer</th>
<th>Hyperlinked Nodes</th>
<th>Backtracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of pages revisited</td>
<td>M = 22</td>
<td>M = 10</td>
</tr>
<tr>
<td>SD = 7</td>
<td>SD = 5.5</td>
<td>SD = 3</td>
</tr>
</tbody>
</table>

**Questionnaire results**

Focusing on the tool’s main goal, which is its ability to link separate information sources available on the web (Q1), the majority of the students (77%) responded with either “Agree” or “Strongly Agree.” Moreover, most of the students (80%) very positively rated the fact that the tool helped them directly access information of interest inside web pages (Q2). These results demonstrated the overall satisfaction with the tool’s goal. Regarding the layer of hyperlinks added over web pages (Q3), most of students’ opinions (77%) were positive, i.e., either “Strongly Agree” or “Agree,” indicating that the new components facilitated navigation through the web pages.

Another encouraging result was the students’ response to the question about whether the tool reduced the amount they needed to remember (Q4). 73% of the opinions were positive, indicating that the tool had reduced the cognitive overhead caused by the navigational learning process. In addition, the great majority of the students (86%) positively evaluated the generated hypertext representation of knowledge by agreeing with the hypothesis that it facilitated quick review of knowledge (Q5). Regarding the ease of use (Q6), only 57% of participants agreed, or strongly agreed, that the tool was easy to use. This result was expected as the way of using traditional web browser was widely known while more effort was required to learn the KnowledgePuzzle’s supporting features.

Students also provided suggestions to enhance the tool’s functionality. For instance, some students suggested the ability to reference not only textual information inside pages but also other components such as videos. Other students suggested the need to share and discuss their personalized navigation paths collaboratively as this will enable them to learn from others’ experiences. Another student suggested the development of the tool as a plug-in to Firefox or Internet Explorer in order to benefit from the advantages of the tool as well as the usability of traditional web browsers.

**Related Work**

The work presented in this article builds upon several areas of research, including information structuring and visualization, adaptive hypermedia and web annotation. The following subsections briefly discuss these areas focusing on their limitation for self-directed learning and navigation planning.

**Information structuring and visualization techniques**

Research in hypertext navigation has proposed several structuring and visualization techniques to help users in avoiding disorientation or in returning to previously visited pages. Cockburn and Greenberg (2000) classified these techniques into four categories: 1. Hub-and-spoke dynamic trees which are generated in response to the user’s navigational acts. 2. Spatial or concept map organizations that aim to exploit people’s memory for the spatial location of objects. 3. Site maps that show a topology of the physical storage locations of pages. 4. Temporal organization schemes that exploit the user’s memory for the timing of their actions. Most of these techniques do not enable user-controlled structuring of information and navigation paths. Only spatial and concept mapping techniques can enable users to structure pages and define relationships based on their own viewpoints.
The design of the KnowledgePuzzle tool was inspired by mind-mapping or concept-mapping techniques which are used to visualize or classify ideas as an aid for organizing information. It has been suggested that these techniques can improve learning efficiency up to 15% over conventional note taking (Farrand, Hussain & Hennessy, 2002). Some systems integrated mind-mapping or concept-mapping approaches into web browsing to help users organize relationships and ideas. Cañas et al. (2005) found that organizing information via a concept map-based interface leads to more accurate search performance than the typically used web page-based browser. Examples of concept-map based browsers include Nestor (Khamidoullina, Reggers & Zeiliger, 2001) and Kashihara et al.’s tool (2000). They often provide a web browser divided into two spaces: a space for web browsing and a space for gathering, representing and structuring information. Other systems demonstrated the ability to create collections of material from disparate websites (e.g., Dontcheva et al., 2006; Schraefel et al., 2002).

The main drawback of the tools described above is the separation between the web browsing process and the planning process, which are often performed in two separate spaces. This separation fragments the reading process and distracts learners as they repeatedly need to move between the browsing space, where they navigate web pages, and the planning (or concept-mapping) space, where they collect and structure information (Chandler & Sweller, 1991; Cockburn & Greenberg, 2000). Although the KnowledgePuzzle tool enables learners to gather and structure information using a concept mapping tool, it makes a step forward to personalize web content by transforming the constructed map into hypertext components and attaching them over web pages. Thus, learners can exclusively focus their attention on the browsing activity which becomes structured and annotated exactly as they need.

Adaptive Hypermedia

Adaptive hypermedia (AH) aims to reduce the cognitive overhead by providing navigational support according to a user model. Having knowledge about the users, disorientation and cognitive overhead can be reduced by organizing, hiding, recommending or annotating links according to the user’s interests (Brusilovsky, 1996). However, AH techniques do not always fulfill the requirements of self-directed learning, where users are free to direct their navigation paths on the web, for the following reasons:

1. AH systems rely on the designer’s predictions of learners’ interests and these may not match their real interests. Although there exist some user-driven techniques that enable users to control the adaptation process (Tsandilas & Schraefel, 2003), this control is still limited to what the designer allows, and it requires specific design settings to be taken before the navigation session.

2. AH techniques can adapt the path of instruction within specific instructional delivery systems. However, the web navigation process may involve web pages from several sites which may use disparate adaptation techniques or may not use any.

3. AH systems are based on user-modeling. User-models are often persistent or change slowly, and their construction is based on assumptions that do not always hold. In self-directed learning, the learner’s desires and goals may change, evolve or propagate as the learning process progresses (Zimmerman, 2002). Thus, an AH system can make incorrect guesses about what the learner wants or it may not be able to capture any shift in the learner’s goals (De Bra, 2000).

In our view, self-directed learning from the web requires a solution that goes further by enabling ordinary Internet users, rather than content authors, to reform the structure of web content by injecting new hyperlinks and annotations in accordance with their way of thinking. We emphasize that what is proposed here is different from link adaptation, the adaptive hypermedia technique. Link adaptation is based on manipulating hyperlinks that already exist in web pages, and thus it offers nothing if the pages of interest are not directly linked in the way that the user wishes.

Web annotation tools

A number of tools have emerged to support the annotation of web pages either manually or semi-automatically. Manual annotation tools allow users to manually add, modify or remove information from an existing web resource without modifying the resource itself. For example, Annotea framework (Kahan et al., 2002) allows users to make and exchange annotations in RDF files. Chatti et al. (2006) use u-Annotea to conduct hand-written annotations on
web pages. Rau et al. (2004) use web annotation tools for learning, allowing learners as well as instructors to annotate, highlight, organize and share annotations on web-based learning material.

Semi-automatic annotation tools such as COHSE (Bechhofer et al., 2008), Magpie (Dzbor, Motta & Domingue, 2004) and Ont-O-Mat (Handscho & Staab, 2002) enable users to inject annotations or browsing links within the web page by integrating ontologies and information extraction tools. They associate an ontology-based semantic layer to web documents, allowing ontological terms within the document to be dynamically annotated or linked to explanatory resources. However, they use static databases for linking and thus they restrict the user’s control over the linking process. Systems like Google’s Autolink also provide dynamic linking functionality.

In contrast to prior annotation techniques, we offer a knowledge-driven linking and annotation service which not only allows learners to annotate web pages but also to build structured associations of web content in order to adapt the navigation path to the learner’s processing of knowledge. Rather than directly interacting with the page content as in manual annotation techniques, learners use a concept-mapping tool to visually plan the navigation path according to their mental representation of knowledge. Subsequently, web content is annotated and interlinked to reflect the constructed visualization.

Conclusion and future work

In order to overcome the differences between what exists on the web and what resides in the learner’s mind, a tool for knowledge construction from the web has been proposed. The tool aims to encourage learners to plan the navigation path and visualize the semantic processing of web content using a concept-mapping tool. Afterwards, the constructed graph is transformed to a layer of informative hyperlinks and annotations, causing the structure of web pages to be adapted to the learner’s information needs. Finally, a hypertext version of the whole knowledge is generated to enable rapid reviewing by enclosing all information assets as well as user-defined associations in a single document. The contribution of the tool is in the proposed approach for self-directed learning that enables learners to operate actively by manipulating, interlinking and annotating static hypertext resources in order to best represent the structure of knowledge in their minds.

In our future work, we will consider the suggestions made by the students to improve the tool. In addition, while the study presented in this article focused on how the tool enhanced the navigational behavior, we will conduct further analysis to explore how the tool can lead to higher knowledge building and acquisition as compared to conventional navigation techniques.

References


Digital Competition Game to Improve Programming Skills

Julián Moreno
GUIAME research group, Universidad Nacional de Colombia, Carrera, Medellín, Colombia // jmoreno1@unal.edu.co

(Submitted March 2, 2011; Revised August 4, 2011; Accepted November 11, 2011)

ABSTRACT
The aim of this paper is to describe a digital game with an educational purpose in the subject of computer programming, which enables students to reinforce and improve their abilities on the concepts of sequencing, defined iteration and nesting. For its design, a problem solving approach was followed and a score comparing mechanism was implemented in order to encourage students to analyze their solutions and look for better ones. To validate the game, a study with a 123 students sample was made, which did not only demonstrate students interest on the approach but also its educational effectiveness.

Keywords
Educational game, Computer programming, Problem solving

Introduction
Gaming, as a learning and skills improvement mechanism, is a natural process not just for humans but also for many other animal species (Huizinga, 1955). From an educational perspective, games have demonstrated to be an auxiliary tool in the construction of students’ systematic knowledge (Lewandowski & Soares, 2008). Systematization through digital games may allow for better students companionship, verifying frequent errors and presenting them multimedia resources in a different and more appealing way compared to traditional classrooms.

A wide variety of studies points out the relevance of these resources to fulfill different learning goals like verbal, mathematical, logic, visual and motor-sensorial, as well as problem solving skills (Klopfer & Yoon, 2005). But, how is that possible? How can digital games help cognitive processes? Piaget (1983) analyzes the importance of the relationship between subject and object (in this case the game), and how object reactions to subject actions allow the latter to build mental models which, through interaction processes, become significant and constitute an important part in learning.

Several works during the last two decades present a general vision about the educational use of digital games, describing its convenience in several areas of knowledge (Cavallari et al., 1992; Randel et al., 1992; Downes, 1999; Duplantis et al., 2002; Rosas et al., 2003; Mitchell & Savill-Smith, 2004; Egenfeldt-Nielsen, 2006; Ke & Grabowski, 2007; Chuang & Cheng, 2009; Kim & Chang, 2010). In the particular case of computer sciences, as it is noted by Becker (2001), the use of educational software is not prominent compared to other disciplines, but new developments have been seen in recent years. Some examples of the application of digital games in different knowledge domains within computer sciences such as software engineering and data structures can now be found in several works (Moser, 1997; Gorriz & Medina 2000; Gander, 2000; Baker et al., 2005; Lawrence, 2004; Papastergiou, 2009; Zapata, 2009).

The work presented on this paper focuses specifically on the improvement of basic programming skills, without emphasis on any particular programming language; i.e., it is related to the algorithms design rather than the final coding. This issue is of great interest because, at least in Colombia, a computer programming class is not mandatory at the high school level, but it is a compulsory part of the curriculum for most engineering related degrees and obviously for technical degrees in computer related fields.

Most teachers in the area would agree that teaching programming logic is challenging because it requires students not only to know how basic instructions work but, more importantly, how to use them in an adequate and creative way to solve a specific problem. For example, if in a classroom with 100 students a teacher proposes a simple exercise like: “Design an algorithm to read three numeric values and determine the largest one”; it would be quite possible for the teacher to find within the group at least 10 different solutions which would not just differ from each
other in an aesthetic way (the use of more or less instructions), but also in, for example, their efficiency regarding the use of resources (processor and memory), which is of great significance from a computational standpoint.

Considering this panorama, an educational digital game called 'ProBot' was designed and developed in our research group with the aim of helping students to reinforce their knowledge on three concepts: sequencing, defined iteration, and nesting. According to Garza (2011), sequencing and iteration are two of the three fundamental elements of computers structured programming (the other one is selection, which is not covered in the game). Sequencing refers to the logical succession provided by the instructions in the algorithm. As they are executed, each instruction in the sequence must logically progress to the next one without producing any undesirable effects. Defined iteration, usually related to the expression \( \text{for} \), is a control structure that refers to the repetition of a set of instructions until certain stopping criterion is met (the number of repetitions is known with antecedence). Once it occurs, the algorithm resumes its logical sequence. Finally, nesting is related to the enclosing of control structures-iteration for example-one into another. It creates layers of instructions where the execution of the interior instructions depends on the execution of the exterior ones.

The game is based on boxing matches where students must fight against several opponents who use certain set of movements. In order to win a match, each student must program his/her own ProBot (abbreviation for Programmable roBot) defining in an algorithmic way the set of movements the ProBot must perform in order to counteract its opponent movements and defeat it. In this sense ProBot is based on the constructionism learning theory proposed by Papert & Harel (1991), which suggests that new knowledge or skills may be acquired more effectively if the learners are engaged in constructing products that are personally meaningful to them. This is done by allowing students to define their own ProBot behaviors in terms that are easy for them to understand and then test them, putting the ProBots in the boxing arena. This approach is also aligned with the constructionism notion suggested by Kafai & Resnick (1996) who state that learners are most likely to become intellectually engaged when they are working on activities that are interesting and significant to them. In this paper, ProBot is described in detail and its performance tested in a real educational environment, proving the power of the computer-aided gaming approach to education and, in particular, to the development of basic algorithmic thinking in students.

**Related works**

There are several works where games are used as learning environments in the computer programming field; some of them are focused in a particular programming language, while others emphasize in basic programming concepts or the development of logic-related skills.

Within the games focusing on a particular language, RoboCode (Li, 2002) and Jupiter (Chicharro et al., 2008) can be found. Robocode is an easy-to-use robotics battle simulator that runs on several platforms supporting Java 2. In this game, instructions on how to manipulate a robot – tank (guns, radar, etc.) must be coded by the player, who then puts it onto a battlefield and let it battle against the opponent robots created by other players in order to see whose code works the best. Jupiter is similar to RoboCode, players may test their Java codes against each other in a competition environment in the form of tournaments, but it is more flexible in the sense that it allows for different kinds of battle games.

Within those games focusing on basic programming concepts there are works like Scratch (Resnick et al., 2009), an educational programming language developed by the Lifelong Kindergarten Group at the MIT that allows people of any background and age (it is intended especially for 6 to 16 year-olds, but it has been used by people in all ages) to experiment with the concepts of computer programming by manipulating visual programming blocks to control some components (images, music and sound). In this category we also have Etoys (Kay, 2005), a child-friendly programming environment based on the idea of programmable virtual entities performing several activities on the computer screen. Both projects, Scratch and Etoys are free and implemented in a language called Squeak (http://www.squeak.org) which is a highly portable, open-source version of Smalltalk.

Finally, there are games like Light-bot (http://armorgames.com/play/2205/light-bot) and some of the games found in http://www.hacker.org which, even if they may not be developed for an educational purpose, incorporate several logic concepts that players must use in order to overcome a set a challenges.
While the game presented here share several similarities with the aforementioned approaches, it also has some fundamental differences. Like RoboCode and Jupiter, one of the main goals of ProBot is to promote the improvement of skills through competence; nevertheless, ProBot is not aimed at experienced students who are already familiar with programming concepts and particular languages; instead, ProBot is similar to Scratch and Etoys in the sense that it is based on graphical programming, where simple blocks representing a basic programming instruction are used instead of pieces of code. Finally, ProBot is similar to Light-bot (in fact, it was after playing this game that the idea of ProBot emerged) and some of the games in hacker.org as, on the one hand, it is also graphical and, on the other, it is based on incremental difficulty challenges.

**Game design**

Considering some elements that promote student involvement into an educational gaming environment (Malone, 1980; Prensky, 2001; Papastergiou, 2009) as well as particular needs of the subject, ProBot was designed following the next considerations, which are included within the elements of this section:

- It has intuitive controls
- It keeps student attention and is encouraging
- It has well defined rules
- It presents immediate and unlimited feedback
- It exhibits step-by-step execution
- It detects student errors and misconceptions
- It has progressive difficulty levels
- As multimedia resource it is appealing and entertaining but without containing extra cognitive load

**Interface**

The game was implemented in Macromedia Flash, so students may access it from a Web navigator after it has been loaded in a server (local or Internet). Its interface is completely visual and is divided in four sections as presented in Figure 1: upper, right, bottom and center.

![Figure 1. Level 1 interface](image)

In the upper section there are three different components (from left to right): the level in which student is currently playing; the available instructions or the corresponding level and the options to enable and disable game sounds and background music. The available instructions are, from a structured programming point of view, the set of structures that are valid for the algorithms a student may define.
In the right section the workspace, where a student can define his/her algorithm, is found. It is here where the student drops in a sequential way the instructions dragged from the upper section: high punch, low block, dodging, etc. Such workspace is analogous to a source code in a programming language, with the difference that the number of ‘code lines’ that may be written is limited by the number of cells in the right section, and each cell can only contain one instruction.

In the bottom section there are two options: in the right, the boxing gloves-like button allows the player to see the “opponent’s training routine,” that is, the set of movements the opponent will perform during the fight; in the left, the bell-like button starts the fight. Once the bell is pressed the ProBot begins to move in a synchronized way with the opponent movements, according to the instructions provided in the workspace. In the central part of the bottom section a small board is found, where key variables of the opponent and the ProBot are shown during the fight. In this board the ‘resistance level’ is presented, as it is common in most fighting digital games, as well as the character’s energy (battery level). The last feature was included into the game looking for students to consider that problems may have resources restrictions.

Finally, the central part of the interface is where action happens. Once a fight starts both, ProBot and opponent, perform their corresponding instructions step by step. If in one of the steps any of the characters receives a punch, its resistance level, which has an initial fixed value in each level, decreases a certain amount. The fight ends when the resistance level of one fighter drops to zero, in which case the other one wins the match. ProBot energy level, which has also an initial fixed value in each level, decreases every time that it punches or dodges, and when this level drops to zero ProBot shut himself down and inevitably loses the match.

**Errors detection**

In the computer programming context, a compiler may be defined as a program that translates a source code, defined in a high level language, into another one in a lower level, typically machine language. This is precisely the main function of the game execution engine in charge of translating the set of instructions the student defines in a set of movements that ProBot performs during the match. A compiler is usually composed of two elements: a lexical analyzer (lexer) and a semantic analyzer (parser). The former validates individual instructions syntaxes whereas the latter validates the global semantic of the source code.

In this game, a lexer is not necessary because individual instructions are not written freely by the student but are chosen from the menu in the upper section of the interface. On the other hand, validating semantic is made by the parser, looking for errors which, in this case, are related to the wrong use of the ‘defined iteration’ concept. In particular, there are two types of errors or misconceptions that students may have during the game: the order of iteration instructions is wrong or the beginnings of the iteration instructions do not match the closures. Both types of errors are detected automatically by the interface when trying to start a fight and informed to the student, so he/she can correct them.

**Difficulty levels**

The game has in total seven difficulty levels that become available as student moves forward. Besides the fact that this is an obvious mechanism for most digital games to maintain the gamer attention, it also has a pedagogical use in the sense that it presents in an incremental order for the concepts to be covered through the game: that is, they are presented in the same order as in the course outline. In particular, the first levels are devoted to the sequencing concept, the following levels add the simple iteration concept, the next ones deal with the concept of multiple iteration and the last ones with the concept of nested iteration. An important feature of each new level is, as presented in Figure 2, the incorporation of additional available instructions as well as more cells in the workspace, allowing the student to define more complex algorithms to solve the corresponding problems.

Each level has a different opponent with his own initial resistance level and punch strength. Each one of them also has its own sounds, gestures and messages; the first two features are included with the aim of entertaining the student, whereas the last one with the aim of encouraging him/her to defeat the opponent. As evident from Figure 1
and 2, the audience increases on each level adding defeated opponents as well as new spectators to the benches in order to incorporate a dramatic effect.

Figure 2. Level 4 interface

<table>
<thead>
<tr>
<th>Player name</th>
<th>Level</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>4</td>
<td>240</td>
</tr>
<tr>
<td>-</td>
<td>4</td>
<td>225</td>
</tr>
<tr>
<td>-</td>
<td>3</td>
<td>220</td>
</tr>
<tr>
<td>-</td>
<td>3</td>
<td>215</td>
</tr>
<tr>
<td>-</td>
<td>4</td>
<td>185</td>
</tr>
<tr>
<td>-</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>-</td>
<td>3</td>
<td>105</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>25</td>
</tr>
</tbody>
</table>

Scoring

As usual in most digital games, a level password element was included so the student can save it to restart the game in other moment from that point once the level has been overcome. Such feature has the additional purpose of identifying the score a student has accumulated along the matches. This is made in an encrypted way to avoid undesired manipulations on such values. In this game, the score is defined as the sum of resistance and energy levels after defeating an opponent, so the accumulated score in level \( n \) corresponds to the sum of the scores obtained in levels \( n, n-1, n-2 \) down to level 1.

As the reader may have inferred already, this score serves to measure the efficiency of the algorithms defined by the student, i.e., the higher the remaining resistance and energy levels after winning a match, the more efficient the corresponding algorithm is. In other words, an opponent can be defeated in many ways, but each way may differ from the others in the resources that are used (energy) and in the punishment that is received in return (resistance). With this feature, good programming practices are promoted through healthy competition. In their search for higher scores, students become aware that, similar to classrooms or real life problems, there are lots of strategies to solve
them, but not all of them are equally efficient. This way, they are forced to think about how “good” their proposed solutions are, and not just about solving the problems.

This phenomenon was validated in the study with actual students described in the next section. It was observed that many students, once they noticed their solutions could be improved after comparing their score with that of fellow classmates, preferred playing again the same level until reaching a higher score, even if they had solved problem (winning the opponent), increasing their skills this way. To allow students to make such comparisons, a web application was built where they could enter their names and level passwords and a dynamic ranking board was displayed with all the group’s scores. Table 1 shows and example of the information that students could see after accessing the web application, where they can clearly measure the efficiency of their solutions as compared to the ones of their classmates. Even though the ‘Last login’ data presented in the second column of Table 1 was not used for comparative purposes during study, it could be used, for example, in fixed-time competitions.

Validation

In order to validate the game proposed in this work, a study was conducted at the Universidad Nacional de Colombia aiming to assess its effectiveness and motivational interest for improving the previously mentioned programming skills on the basis of the curricular objectives pertaining to a course in engineering. The sample was composed by 123 students attending the course ‘Programming fundamentals,’ aged 16–18 years old, 74 males, 49 females, all of them with basic computer knowledge, including Web browsing (to access the game). The sample was divided randomly into two groups: a control group and a test group.

To determine if there was a significant difference among the general previous performance of the two groups, a hypothesis test for independent samples via the t statistic was applied as well as an effect size analysis via the Cohen’s $d$. In both cases the data from the first exam of the course were used. Such an exam did not cover the three concepts of interest: sequencing, iteration and nesting; but related concepts like the logical operators and the binary numeric system. For this reason such an exam cannot be considered as a pre-test per se, but as a measure of previous related performance of the two groups. The corresponding summary data is presented in Table 2, where grades are on a 0-5 incremental scale with a decimal number, as usual in most Colombian higher education institutions. The null hypothesis ($H_0$) was that mean grades were equal and consequently the alternative hypothesis ($H_1$) was that they were not.

<table>
<thead>
<tr>
<th>Group</th>
<th>Students number</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>63</td>
<td>3.206</td>
<td>0.674</td>
</tr>
<tr>
<td>Test group</td>
<td>60</td>
<td>3.315</td>
<td>0.776</td>
</tr>
</tbody>
</table>

The value of the $t$-statistic is -0.830 with a corresponding $P$-value of .408. As the $P$-value is higher than .05 it could be said, with a 95% significance level, that the null hypothesis is accepted. In other words, there is not statistical evidence of previous difference between both groups regarding their performance in related concepts. Consequent with this result, the Cohen’s $d$ value was 0.109 which implies a small effect size.

After teaching to both groups all required concepts in a regular classroom according to the course syllabus during four weeks, a regular set of exercises was given as homework for two weeks to the students of both groups, while students from the test group, apart from the regular exercises interacted with the game during same period of time. After that time, a test was made simultaneously to students of the two groups containing four problems that covered all of the taught concepts. The results of the test are summarized in Table 3.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>3.471</td>
<td>0.579</td>
</tr>
<tr>
<td>Test group</td>
<td>3.847</td>
<td>0.581</td>
</tr>
</tbody>
</table>

To determine if there was a significant difference among the results of the two groups, a hypothesis test was performed again. In this case, the value of the t-statistic using data from Table 3 is -3.594 with a corresponding $P$-
value < .001. According to this value the null hypothesis is rejected in favor of the alternative hypothesis, that is, the mean grades are not equal between the groups, with a difference of 0.376 (10.8%) in favor of the Test group. Additional to this result, the Cohen’s $d$ value was in this case 0.654, implying a medium effect size.

Although this study demonstrated game effectiveness to improve programming skills, means difference was not as large as initially expected, so a more detailed analysis was made. This time, instead of analyzing general test results, the individual results per exercise were analyzed, considering that each one corresponds to a specific concept. Summary of these results is presented in Table 4.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Control group Mean</th>
<th>Test group Mean</th>
<th>Control group Standard deviation</th>
<th>Test group Standard deviation</th>
<th>$t$-statistic</th>
<th>$P$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing</td>
<td>4.100</td>
<td>4.065</td>
<td>0.559</td>
<td>0.582</td>
<td>0.340</td>
<td>.735</td>
</tr>
<tr>
<td>Simple iteration</td>
<td>4.021</td>
<td>4.042</td>
<td>0.607</td>
<td>0.567</td>
<td>-0.198</td>
<td>.843</td>
</tr>
<tr>
<td>Multiple iteration</td>
<td>3.252</td>
<td>3.795</td>
<td>0.599</td>
<td>0.719</td>
<td>-4.539</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Nested iteration</td>
<td>2.451</td>
<td>3.447</td>
<td>1.088</td>
<td>0.820</td>
<td>-5.751</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

In the case of the more basic concepts, sequencing and simple iteration, the $P$-values determine that such differences are not statistically significant. In the same way, the Cohen’s $d$ values (-0.062 and 0.036 respectively) determine a small effect size. A completely different panorama appears when analyzing the more complex concepts, multiple and nested iteration, where the $P$-values determine a significant difference in favor of the test group. In these cases the Cohen’s $d$ values (0.829 and 1.039 respectively) determine a large effect size.

Besides previously described analyses, and with the aim of measuring students’ feeling about the game, a questionnaire with the next five questions was conducted in the control group:
1. What is your overall appreciation about the game?
2. Was it useful to reinforce sequencing, defined iteration, and nesting concepts?
3. How much did you like the graphics, animations, sounds, and effects?
4. Were the interface and controls easy to understand?
5. Was the levels’ difficulty appropriate?

All questions could be answered on a Likert scale from zero to five using only integer values. For questions 1 and 3 the students’ answers refer to a perceptive score (being zero the lowest and five the highest), whereas for the remaining three they could be interpreted as a conformity measure from zero = strongly disagree to 5 = strongly agree. The results for the 60 students are summarized in Table 5.

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.383</td>
<td>0.666</td>
</tr>
<tr>
<td>2</td>
<td>4.017</td>
<td>0.873</td>
</tr>
<tr>
<td>3</td>
<td>3.700</td>
<td>0.788</td>
</tr>
<tr>
<td>4</td>
<td>4.133</td>
<td>0.769</td>
</tr>
<tr>
<td>5</td>
<td>3.867</td>
<td>0.833</td>
</tr>
</tbody>
</table>

From table 5, it is clear that the general perception about the game was quite positive and most students found it easy to interact with (questions 1 and 4). They also seemed to be satisfied with the game as a tool in their learning process (question 2). As one of the students in the test group said, ‘It is good to learn and practice doing something that is not boring.’ The question with highest standard deviation was question 5 because some of them found the increasing level of difficulty adequate, whereas in others opinion ‘Last levels are too difficult.’ The question with lower score was question 3 and, when asking them about it, most said that even though the multimedia aspects were fine, they are used to more sophisticated and realistic 3D games of commercial consoles.

Another important issue about students’ perception is that, after answering the questionnaire, when a discussion started and students could talk freely about their personal experience, many of them agreed the scoring and particularly the comparison mechanism presented in table 1 was an important encouraging instrument that forced
them to redefine their solutions and therefore improve their skills. ‘Every time I finished a level, I checked the scoring table to see how good I was compared to my classmates,’ another student said.

After this experiment, other analyses were performed but this time considering the students’ results in the final exam of the course at the end of semester, six weeks after the previously described test. The results from that exam, considering only the questions related to the concepts of interest, are presented in Table 6.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>3.251</td>
<td>0.498</td>
</tr>
<tr>
<td>Test group</td>
<td>3.642</td>
<td>0.471</td>
</tr>
</tbody>
</table>

With data from Table 3 and Table 6, two comparisons were done, using again hypothesis tests and Cohen’s $d$ values. The first comparison consisted in determining if the performance difference between both groups was as significant during the final exam as it was during test. The value of the $t$-statistic was -4.475 with a corresponding $P$-value < .001. This means that there is statistical evidence of better performance of the students that interacted with game (a difference of 0.391 or 12%) even in the final exam. In addition, the Cohen’s $d$ value in this case was 0.813 implying a large effect size.

In the second comparison, and with the aim of determining how the students’ retention about the concepts of interest was impacted, I contrasted both groups performance during test and final exam. For the test group the latter grades were lower (the mean grade drops 0.205 or 5.3%) and the $t$-statistic was 2.154 with a corresponding $P$-value of .033. The corresponding Cohen’s $d$ value was 0.390, indicating a medium effect size. For the control group where the fall of the mean grade was larger (0.220 or 6.3%) and the panorama was similar: the $t$-statistic was 2.263 with a corresponding $P$-value of .025, and the Cohen’s $d$ value was 0.410. These results imply that there is no statistical evidence of the students’ retention, neither for the ones who used the game, nor for the ones who do not.

Conclusions

Use of digital games in classrooms with educational purposes has been increasing during the last decades and there are numerous studies validating its effectiveness in the learning process. Considering this scenario, the work described in this paper presents a contribution in the specific domain of computer programming, with a digital competition game to improve skills related to sequencing, defined iteration and nesting.

To validate the game, a study with 123 students from a basic course for engineering degrees in a Colombian university was conducted. Such a study demonstrated that the game was not only appealing for most of them, but also that the students who used it achieved a better understanding of the concepts of interest compared to those who did not. These results coincide with the findings of other contributions on the pedagogical use of games, especially because of the motivational aspect (Lucas et al., 1975; Garris et al., 2002; Barnes et al., 2007; Wilson et al., 2009). While such phenomenon was validated in a general way, it was more evident when the most complex concepts were considered separately.

Some of the features that allow such a success, helping students to reinforce prior knowledge and to improve the corresponding skills, were the progressive difficulty levels, errors detection, step-by-step execution and immediate feedback. Other main features were the scoring mechanism and the scores board which provided a differential factor, encouraging students to review and improve their solutions.

What this study did not prove was the game impact on the student’s retention. Although in two different evaluations, with a difference of six weeks between, the students of the test group obtained better results than those of the control group, there was a fall in the mean grades for both groups during the latter evaluation. In other words, even if the game helps students to reinforce the concepts of interest, it seems those who use it, do not have an extra advantage for escaping from the forgetting process, compared to those who do not.

As a final conclusion one may say that from the multimedia point of view, even though most of the students confessed they are familiar with more sophisticated games, the game described here was still found attractive and
entertaining. While this issue is not necessarily a main goal for educational purposes, it is in fact very relevant to capture students’ attention which, as many teachers know, is a really difficult task especially in teenagers and young people like the ones who participated in the study.

As future work we are already planning to develop new versions of ProBot taking advantage of what we have found so far. Some of the ideas include adding extra difficulty levels, its extension to other kinds of competitions and sports, and the incorporation of some other basic concepts related to computer programming, like conditionals and recursion. We would also want to perform a larger study about the game, analyzing other important issues like how gender, high school background or home computer use affect its effectiveness; or how can it help students in lower grades to improve the skills related to the considered concepts.

Readers interested in trying ProBot in its current version (without the scoring table mechanism) for educational purposes or just for having fun can access it from http://pisis.unalmed.edu.co/probot/.

References


A Comparison of Demonstration and Tutorials in Photo Editing Instruction

Cenk Akbiyik
Erciyes University, Faculty of Education, Kayseri, Turkey // cakbiyik@erciyes.edu.tr

(Submitted March 14, 2011; Revised July 7, 2011; Accepted September 19, 2011)

ABSTRACT
Throughout this study, tutorial and demonstration methods in photo editing instruction were compared. The instruction method was the independent variable of the study while cognitive load, state anxiety level, student performance, student achievement, and task duration were dependent variables. The research was conducted with 62 undergraduate students. The students in the tutorial group fulfilled the tasks following tutorials. Those in the demonstration group were first given a demonstration, and then were asked to perform the tasks. Results indicated that students from the demonstration group showed significantly better task performance and had significantly lower state anxiety levels than did the students in the tutorial group. Both groups were found to be equal in regard to achievement test scores, recall test scores, and task duration. Students tended to feel more anxious during the more difficult tasks. Both cognitive load and state anxiety levels were found to have lengthening effects on task duration. Both methods have their own strengths and weaknesses. The most important point in both methods emerged as instructor support.

Keywords
Computer instruction, Educational technology, Cognitive load, State anxiety, Demonstration, Tutorial

Introduction
A very significant point in the teaching-learning process is the design of teaching-learning experiences. Scientists, educators, and instructional designers carry out vast numbers of studies to enable them to answer questions such as “What kind of learning environment should be designed?”, “Which strategies and methods should be employed?”, “Which educational materials should be used?”, and “What should the role of the instructor be?”

In response to economic and social change, countries all over the world are formulating policies that incorporate the use of information and communications technologies (ICT) in education (Vanderlinde, & van Braak, 2011). There is a widespread belief that computer skills are and will continue to be increasingly necessary for every individual to participate effectively in society, the economy, and the workforce. In all likelihood our world will increasingly require the use of computers and computer-related technologies. Therefore, effective and efficient instruction in computer use is highly important. Unfortunately, there is little research or documentation on the teaching strategies employed in end-user computer education and training (Phelps, Hase, & Ellis, 2005). Indeed a current literature search on computer instruction shows an insufficient number of results. Clearly more studies are needed on the planning and conducting of computer instruction.

It is crucial to examine how we can benefit from learning environments and how we can design learning environments to maximize student learning. Cognitive load theory is one of the theories that deals with the development of effective instructional materials and learning environments. Basically it grounds itself on visual and auditory information channels that are partially independent from each other (Kilic, 2006). According to the cognitive load theory, eliminating excessive load on auditory and visual channels is crucial so that information can be transferred to the long-term memory (Sweller, 1994).

The cognitive load theory states that human working memory is limited and that it can deal with only two or three elements at a time. This limited type of memory may be affected by three types of load: intrinsic, extraneous, and germane. The intrinsic cognitive load cannot be reduced or modified by instructional design because it involves the difficulty of the content. Extraneous cognitive load derives from instructional materials that are used to present information. This type of cognitive load is generated by the instructional format used. High extraneous cognitive load occurs when the instructional material is poorly designed. Changing instructional materials to facilitate learning is one way to reduce extraneous cognitive load (Adulserane, 2007). When both intrinsic and extraneous cognitive loads are high, the working memory will be overloaded. Cognitive overload may be defined as the confusion of...
learners because of too many options, components, and ways (Murray, 2001). A decrease in learner performance is expected during high cognitive load (Paas, Renkl & Sweller, 2004).

Anxiety is a pervasive feeling of uneasiness based on real or imaginary stimuli that can produce secondary reactions, either physiological or psychological, or a tertiary reaction combining both the psyche and biological domains (Fisher, 2008, p. 18). Although multiple definitions have been introduced for anxiety and to test anxiety, there is broad agreement that anxiety can be classified into two components, state anxiety and trait anxiety. It is important to understand these concepts because one can delineate differences in state and trait anxiety. State anxiety can be altered because it refers to how a person copes with and responds to different levels of anxiety and stress in his/her life (Prato, 2009, p. 10). Trait anxiety is unchangeable and is how an individual reacts to anxiety on a day-to-day basis. These two kinds of anxieties are not independent from each other. They have positive correlations ranging between 0.33 and 0.66 (Basarir, 1990, cited in Duman, 2008).

Research shows academic achievement has a correlation with emotional traits (Yildirim, 2000). State anxiety also has a correlation with learning performance. The well-known Yerkes-Dodson Law (Smith, Sarason, & Sarason, 1982, cited in Macintyre, 1995), for example, describes a curvilinear relationship between anxiety and performance as a function of task difficulty. When a given task is relatively simple, anxiety seems to have little negative effect and may actually improve performance through increased mental effort. However, as the demands on the system increase, the extra effort may not fully compensate for the cognitive interference, and anxiety will begin to have a negative effect.

We will not be mistaken if we state demonstration and tutorials are two ways widely used in computer instruction. Unfortunately, a literature search shows these ways were not compared by means of two important variables, cognitive load and state anxiety. The aim of this study was to compare demonstration and the use of tutorials in photo editing instruction. It was assumed that this study would address a real need and make valuable contributions to selecting optimal methods in instruction of software applications. Because the process of learning how to use computers should be examined from multidisciplinary perspectives (Yan & Fischer, 2004), within the content of the study, two important variables, cognitive load and state anxiety levels, were taken into consideration as well as certain other variables such as student performance, student achievement, and task duration.

Method

This research is an experimental study. The independent variable of the study is a teaching method that has two categories: use of tutorials and demonstration. There are several dependent variables in this study. However, the study has similar aspects with action research. Action researchers focus on situations that may enable them to change certain conditions in which they are involved (Buyukozturk et al., 2009). In this study, the researcher himself is an experienced ICT instructor who has participated actively in the process. In addition, the researcher prepared instructional materials (tutorials) and collected data.

This study investigates the following research questions:
1. Are there significant differences between the groups regarding task performance scores, achievement scores, recall test scores, task duration, cognitive load levels, and state anxiety levels?
2. What are the cognitive load levels, state anxiety levels, and task duration of the treatment groups throughout the study?
3. How are the cognitive load levels, state anxiety levels, task performance scores, achievement scores, recall test scores, and task duration of the groups related?
4. What are the differences between the study groups in the way they fulfill tasks?
5. What are the opinions of the participants about the tutorial and demonstration methods employed in photo editing instruction?

Participants

Sixty-two undergraduate students enrolled in the Turkish studies program in the faculty of education at Erciyes University participated in the study. The participants took a computer literacy course (Computer I) before the
research began and were in the process of taking Computer II course during the study. The participants were attending the course in two separate groups. These groups were assigned randomly to tutorial and demonstration conditions. Thirty students were in the tutorial group and 32 in the demonstration group.

Most of the participants were between 19 and 22 years of age. The experimental groups were identical with regard to gender distribution (Table 1), and computer literacy course academic achievement \( t = -0.98, p < 0.1 \) (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Tutorial</th>
<th>Demonstration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>f</td>
<td>14</td>
</tr>
<tr>
<td>%</td>
<td>22.6%</td>
<td>22.6%</td>
</tr>
<tr>
<td>Male</td>
<td>f</td>
<td>16</td>
</tr>
<tr>
<td>%</td>
<td>25.8%</td>
<td>29.0%</td>
</tr>
</tbody>
</table>

### Table 2. Comparison of computer literacy achievements

<table>
<thead>
<tr>
<th>Group</th>
<th>( N )</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>32</td>
<td>84.02</td>
<td>19.74</td>
<td>-0.98</td>
<td>0.33</td>
</tr>
<tr>
<td>Group 2</td>
<td>32</td>
<td>88.22</td>
<td>14.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Implementation

The study was implemented during the Computer II course and lasted for eight lessons. In the first six lessons, students were assigned a different photo editing task each day. The students in the tutorial group fulfilled the tasks following tutorials. The demonstration group was first given a demonstration, and then were asked to perform the tasks. Students fulfilled all the assignments using the same photo editing software. All participants recorded their screen images while they were performing the tasks. After each task they saved the task files and responded to the cognitive load and state anxiety level surveys. The following class, students took the achievement test. In the last class, students responded to the recall test and student opinions survey. In addition, the duration of each demonstration was recorded by the instructor.

### Materials

Action research relies chiefly on observation and behavioural data (Cohen & Manion, 1994). In this study, all educational materials were developed by the researcher after a literature search and observation of teacher candidates’ needs over a period of two years. Both the assignments and the tutorials were prepared within the borders of the following learning objectives:

1. Using selection tools (lasso, magnetic lasso, rectangular marquee, elliptical marquee)
2. Inversing a selection
3. Setting “feather” property of selection tools
4. Copying and pasting selected parts of an image
5. Using paint bucket, brush, and clone tools
6. Changing brush type and size
7. Moving, resizing, and rotating parts of an image
8. Adding, deleting, and ordering layers
9. Adjusting the opacity level of a layer
10. Adding text to an image
11. Changing colour balance, saturation, and brightness levels of an image
12. Adding filters to an image

The same photo editing assignments, selected from a pool of photo editing assignments, were given to both of the groups. Before being chosen for this study, all assignments in the pool were tested (in previous years) and corrected.
The tutorials prepared for the tutorial group were also developed by the researcher after a literature review. A step-by-step explanation style was preferred in the tutorials. They were tested and corrected before this research.

**Data collection**

Data were collected at the beginning of the spring semester of 2010. The data collection process took place in the computer labs where students took their computer courses. Data were collected through various tools such as the cognitive load level survey, the state anxiety level survey, student assignment files, videos of student assignments, student achievement test, recall test, and student opinions survey.

**Data collection tools**

Cognitive load levels were determined by the cognitive load survey. The survey was taken from Kilic (2006), who developed and used the Turkish version. Students scored a perceived cognitive load level on a scale of 1 to 5 after the completion of each task.

State anxiety levels were determined by the intensity of emotions survey. This survey was adapted from the work of Serbest (2007). Intensity levels for four positive and eighteen negative emotions were asked for in the survey. Students scored perceived intensity of emotions on a 1 to 5 point scale after completion of each task. Within the scope of this study only state anxiety levels were taken into consideration.

In this study, student performance is indicated by the scores of the six photo editing tasks that students fulfilled during the lessons. These six tasks were completed by the groups two different ways, by demonstration and by tutorial. Students could re-apply to the tutorial (tutorial group), ask to the instructor questions (demonstration group), and ask each other questions (both groups) while fulfilling these tasks. Performance scores were determined with the aid of a rubric. The rubric was constructed from the photo editing operations that students were asked to fulfil. Each operation was counted as one point.

Achievement scores were determined by the assessment of the two photo editing tasks that students were given after the lessons. Students were given neither a tutorial nor a demonstration while they were fulfilling these two tasks. In addition, they were not allowed to ask each other any questions. Achievement points were also determined with the aid of a rubric, which was made up of the photo editing operations.

For the recall test, students were asked to write what and how much they remembered from each of the six photo editing assignments. The rubric given in Table 3 was used to determine the recall test scores.

<table>
<thead>
<tr>
<th>Points</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Concerning the assignment, the student remembered nothing at all, or remembered nothing correctly.</td>
</tr>
<tr>
<td>1</td>
<td>Concerning the assignment, the student remembered some of the operations correctly, but did not give any explanation.</td>
</tr>
<tr>
<td>2</td>
<td>Concerning the assignment, the student remembered some of the operations correctly and explained them.</td>
</tr>
<tr>
<td>3</td>
<td>Concerning the assignment, the student remembered all of the operations correctly and explained them briefly.</td>
</tr>
<tr>
<td>4</td>
<td>Concerning the assignment, the student remembered all of the operations correctly and explained them in detail.</td>
</tr>
</tbody>
</table>

Videos of student assignments were screen recordings that were recorded by the students during each assignment. All videos were watched and evaluated qualitatively by the researcher. In addition, task duration was determined from these videos, but the duration of each demonstration was added onto the demonstration group’s task duration.

The student opinions survey consisted of two open-ended questions. Students were asked to assess the method they encountered during the photo editing instruction. They were also asked to assess the other method they did not encounter. Content analysis was used to analyze the students’ responses qualitatively.
Data analysis
Throughout the research, frequency calculations, percentage calculations, line charts, Mann-Whitney U tests, and Pearson correlation coefficients were used for the quantitative analyses. Content analysis was used for the qualitative analyses.

Results
The first research question of the study was about the comparison of the two groups concerning task performance scores, achievement test scores, recall test scores, task durations, cognitive load levels, and anxiety levels. Because group sizes were not adequate to perform a parametric test, several Mann-Whitney U tests were used to make the necessary comparisons. The results illustrated in Table 4 show us that there are significant differences between the two groups regarding task performance scores and state anxiety levels ($p < 0.1$). These differences are in favor of the demonstration group. In other words, the students in the demonstration group have higher performance scores and lower state anxiety levels. The groups are identical in terms of achievement test scores, recall test scores, task durations and cognitive load levels.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Task performance score</th>
<th>Achievement test score</th>
<th>Recall test score</th>
<th>Task duration (seconds)</th>
<th>Cognitive load level</th>
<th>State anxiety level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorial</td>
<td>30</td>
<td>8.21</td>
<td>22.65</td>
<td>30</td>
<td>3.41</td>
<td>2.03</td>
</tr>
<tr>
<td>Demonstration</td>
<td>32</td>
<td>8.99</td>
<td>39.80</td>
<td>27</td>
<td>2.92</td>
<td>1.78</td>
</tr>
</tbody>
</table>

Cognitive load levels, anxiety levels, and task duration throughout the study are given in Table 5. In order to make a better representation of the data given in Table 5, bar charts were drawn for each of the three variables.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
<th>Task 5</th>
<th>Task 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive load levels</td>
<td>Tutorial</td>
<td>3.41</td>
<td>3.56</td>
<td>3.76</td>
<td>3.13</td>
<td>2.86</td>
</tr>
<tr>
<td>State anxiety levels</td>
<td>Tutorial</td>
<td>2.30</td>
<td>2.38</td>
<td>1.88</td>
<td>1.63</td>
<td>1.92</td>
</tr>
<tr>
<td>Task duration</td>
<td>Tutorial</td>
<td>763.00</td>
<td>739.00</td>
<td>706.00</td>
<td>455.00</td>
<td>350.00</td>
</tr>
</tbody>
</table>

Chart 1 illustrates the cognitive load levels for the two groups throughout the study. It shows that the cognitive load levels for the tutorial group experience a slightly upward trend for the first three tasks. But then, the levels begin to decline and students complete the last task with a lower cognitive load level than the initial one. On the other hand, cognitive load levels for the demonstration group have slight falls and rises throughout the tasks.

Chart 2 illustrates the state anxiety levels for the two groups. Here we see that the anxiety levels for both groups decline for the first four tasks. The anxiety levels continue to fall for the demonstration group while they rise slightly for the tutorial group.
Chart 1. Cognitive load levels throughout the study

Chart 2. State anxiety levels throughout the study

Chart 3 represents the task durations (in seconds) for both of the groups. From the chart we can see parallel task duration for both groups. Although there is a downward trend for the first five tasks, the duration suddenly increases for the last task.

Chart 3. Task durations throughout the study

Table 6 shows the relationships between cognitive load level, state anxiety level, task performance scores, achievement test scores, and recall test scores. From the results we can see that the correlations between the variables are not completely parallel for both groups, instead some variations can be seen. Cognitive load levels and anxiety levels have moderate or weak positive correlations with each other. In other words, students tend to feel more anxious during more difficult tasks.
Task performance scores have varying relations with cognitive load levels and state anxiety levels. For the tutorial group, task performance has weak positive correlations with cognitive load and anxiety. However, for the demonstration group, task performance is correlated negatively with cognitive load and has no correlation with anxiety levels.

Table 6. Relations between cognitive load level, state anxiety level and task performance score, achievement test score, and recall test score

<table>
<thead>
<tr>
<th>Cognitive load level</th>
<th>State anxiety level</th>
<th>Task performance score</th>
<th>Achievement test score</th>
<th>Recall test score</th>
<th>Task duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorial</td>
<td>1.00</td>
<td>0.25</td>
<td>0.23</td>
<td>0.00</td>
<td>0.29</td>
</tr>
<tr>
<td>Demonstration</td>
<td>1.00</td>
<td>0.47</td>
<td>-0.21</td>
<td>-0.37</td>
<td>-0.13</td>
</tr>
<tr>
<td>Cognitive load</td>
<td>0.25</td>
<td>1.00</td>
<td>0.20</td>
<td>-0.07</td>
<td>0.30</td>
</tr>
<tr>
<td>State anxiety level</td>
<td>0.47</td>
<td>1.00</td>
<td>0.05</td>
<td>-0.41</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Achievement test scores show no correlation with regard to the cognitive load level and anxiety level for the tutorial group. However, for the demonstration group the scores have moderate negative correlations with cognitive load and anxiety levels.

Recall test scores also have varying relations with cognitive load levels and anxiety levels. For the tutorial group, recall test scores show moderate levels of positive correlations with cognitive load and anxiety. However, for the demonstration group, the scores have a weak negative correlation with cognitive load and a weak positive correlation with anxiety levels.

Finally, correlations between task duration and cognitive load and anxiety levels are all positive and weak, except one. From the results, we can see that there is a moderate level of positive correlation between task duration and anxiety level for the demonstration group.

The fourth research question was formulated as “What are the differences between the groups in fulfilling the tasks?” To answer this research question, screen recordings of the participants were viewed and evaluated quantitatively. One hundred and forty videos from the tutorial group and 134 videos from the demonstration group were examined. The following codes were formulated by the researcher during the analyzing process. Codes and their frequencies are given in Table 7.

Table 7. Codes and code frequencies for video analysis

<table>
<thead>
<tr>
<th>Code</th>
<th>Tutorial</th>
<th>Demonstration</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Following instructions or explanations</td>
<td>29</td>
<td>14</td>
<td>43</td>
</tr>
<tr>
<td>Making mistakes</td>
<td>20</td>
<td>14</td>
<td>34</td>
</tr>
<tr>
<td>Fulfilling rapidly</td>
<td>12</td>
<td>17</td>
<td>29</td>
</tr>
<tr>
<td>Fulfilling correctly</td>
<td>11</td>
<td>18</td>
<td>29</td>
</tr>
<tr>
<td>Making unnecessary moves</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Trial and error</td>
<td>12</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Confusion</td>
<td>5</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Fulfilling without any difficulty</td>
<td>3</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Referring back to the instructions</td>
<td>7</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Omitting steps</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

We can see that while fulfilling the tasks, students from both groups preferred to follow the instructions or explanations instead of learn by trial and error. However, students from the tutorial group applied to the trial and error way more than the others. When we examine the frequencies of positive codes such as “fulfilling rapidly” or “fulfilling correctly,” we can see that positive codes have higher frequencies in the demonstration group. On the other hand, the demonstration group also has higher frequencies for negative codes such as “making unnecessary moves” or “confusion.” This result shows us that some students from the demonstration group fulfilled the tasks correctly without confusion or difficulty. Other students from the same group missed or forgot instructions, resulting in some difficulties in fulfilling the tasks. When we compare frequencies of making mistakes for both groups, it can be seen that students from the tutorial group made more mistakes than other students.
Student opinions were collected with two open-ended questions. Responses were analyzed using content analysis. Tables 8 and 9 contain the codes, code frequencies, and thematic codes of this analysis.

Table 8. Content analysis of students’ opinions (tutorial group)

<table>
<thead>
<tr>
<th>Codes</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>We had the opportunity to re-apply to the steps in the event of forgetting.</td>
<td>1</td>
</tr>
<tr>
<td>Steps do not get confused with each other.</td>
<td>1</td>
</tr>
<tr>
<td>Tutorials helped me to see the steps better.</td>
<td>1</td>
</tr>
<tr>
<td>It is good to have all steps in front of you.</td>
<td>3</td>
</tr>
<tr>
<td>I could not recall previous tasks.</td>
<td>1</td>
</tr>
<tr>
<td>It was easy to understand and apply the steps. Because they were designed from simple to complex.</td>
<td>1</td>
</tr>
<tr>
<td>Tutorials enabled me to do the tasks on my own effort.</td>
<td>2</td>
</tr>
<tr>
<td>Tutorials enabled me to do the tasks without anybody’s help.</td>
<td>1</td>
</tr>
<tr>
<td>Tutorials helped me to discover my skills.</td>
<td>1</td>
</tr>
<tr>
<td>Demonstration of some steps by the instructor is more effective.</td>
<td>3</td>
</tr>
<tr>
<td>Students should study with tutorials and consult the instructor when they need to.</td>
<td>3</td>
</tr>
<tr>
<td>Instructor should first make some explanations and then ask students to use a tutorial.</td>
<td>1</td>
</tr>
<tr>
<td>Tutorials are fun.</td>
<td>1</td>
</tr>
<tr>
<td>Tutorials are effective.</td>
<td>8</td>
</tr>
<tr>
<td>I performed the steps very slowly.</td>
<td>1</td>
</tr>
<tr>
<td>I achieved the goals more quickly.</td>
<td>1</td>
</tr>
<tr>
<td>It is easier to do the tasks with tutorials.</td>
<td>1</td>
</tr>
<tr>
<td>I had difficulties studying with tutorials.</td>
<td>1</td>
</tr>
<tr>
<td>It is more difficult to understand the tutorials and perform the steps.</td>
<td>1</td>
</tr>
<tr>
<td>Tutorials had a negative effect on my performance.</td>
<td>1</td>
</tr>
<tr>
<td>I may not remember all the steps if I learn by demonstration.</td>
<td>2</td>
</tr>
<tr>
<td>It is easier to do the steps if someone demonstrates.</td>
<td>3</td>
</tr>
<tr>
<td>Retention of knowledge is low when we learn with demonstration.</td>
<td>2</td>
</tr>
<tr>
<td>We can ask a lot of questions if we learn by demonstration.</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thematic codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity of the steps in the tutorials</td>
</tr>
<tr>
<td>Ease of access to all steps</td>
</tr>
<tr>
<td>Opportunity of reapplication to steps</td>
</tr>
<tr>
<td>Opportunity for individual study</td>
</tr>
<tr>
<td>Decreasing need for instructor</td>
</tr>
<tr>
<td>Need for help and guidance of instructor in some cases</td>
</tr>
</tbody>
</table>

From Table 8, we can understand that tutorials provided the students with the opportunity for individual study. Tutorials made access to instructions easier. Although some students had difficulties with the tutorials, they were perceived as positive and effective by most of the students. In general, students tended to think that tutorials decreased but not completely diminished the need for an instructor. While some students were happy with solving problems by themselves, some preferred to consult the instructor if necessary. Moreover, some students from this group expressed the view that they could learn more with the explanations of an instructor.

Table 9. Content analysis of students’ opinions (demonstration group)

<table>
<thead>
<tr>
<th>Codes</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstration makes tasks easier to understand.</td>
<td>3</td>
</tr>
<tr>
<td>We may miss some explanations.</td>
<td>1</td>
</tr>
<tr>
<td>We ask each other if we miss something, enabling us to repeat the steps.</td>
<td>1</td>
</tr>
<tr>
<td>Demonstration is effective.</td>
<td>10</td>
</tr>
<tr>
<td>We realize our mistakes with demonstration.</td>
<td>1</td>
</tr>
<tr>
<td>Demonstration has a positive effect on learners.</td>
<td>1</td>
</tr>
<tr>
<td>Demonstration makes tasks easier.</td>
<td>1</td>
</tr>
<tr>
<td>We can ask the instructor questions if we need to.</td>
<td>2</td>
</tr>
<tr>
<td>Demonstration makes asking the instructor questions easier.</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thematic codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of comprehension of instructions</td>
</tr>
<tr>
<td>Probability of missing some instructions</td>
</tr>
<tr>
<td>Peer support in demonstration</td>
</tr>
<tr>
<td>Positive student perceptions towards demonstration</td>
</tr>
<tr>
<td>Opportunity to get help and guidance from an instructor</td>
</tr>
</tbody>
</table>
Both ways are effective.  
Face to face learning is more effective.  
Demonstration affects retention more.  
Demonstration is better than using tutorials.  
Using demonstration together with tutorials could be more effective.  
I do not think tutorials are effective.  
We may have difficulties if we study with tutorials.  
Studying with tutorials may decrease social interaction.  
No feedback from the instructor may have negative effect on our performance.  
We may have difficulties in performing the steps about which we do not have any preliminary knowledge.  
Tutorials may be helpful for complex tasks.  
It would be more effective if we use tutorials whenever we need to.  
I would be more confident if I had learnt with tutorials.  
Tutorials may have negative effects on motivation.  

Idea of effectiveness of both ways  
Idea that demonstration is more effective than tutorials  
Student perceptions towards ineffectiveness of tutorials  
Student perceptions towards higher difficulty of lessons with tutorials  
Student perceptions towards negative effects of tutorials on social interaction  
Idea of students may apply to tutorials if they need  
Idea of tutorials may have negative effects on student motivation  
Idea of tutorials may have negative effects on student self-confidence

It could be difficult to find answers from tutorials if I was in difficulty.  
Using tutorials means that the instructor took the easy way.  
Instructors should not demand any knowledge they have not explained.  

Student perception towards instructor is taking the easy way if he is using tutorials

From the table we can see that students have positive perceptions towards the demonstration method. This positive perception was due to the ease of comprehension of the instructions, instructor support, peer support, and increased social interaction. However, some students stressed that missing or forgetting instructions as negative sides of the demonstration method. Students from this group generally had negative perceptions towards the use of tutorials. They expressed the possible negative effects of tutorials on student motivation and difficulty of tasks. Moreover, some of the students considered the instructor’s use of tutorials as a sign of the instructor taking the easy way.

**Discussion**

The first finding of the study was the existence of significant differences between the two groups regarding task performance scores and state anxiety levels. Students from the demonstration group showed a significantly better task performance. This finding was supported with the result of assignment video analysis. Students from the tutorial group made more mistakes than did the others. Furthermore, the demonstration group had lower state anxiety levels. These findings may be explained by instructor support. The demonstration group’s students could freely consult the instructor during lessons. This support may have helped the students to perform better and to feel less anxious. As Yan & Fischer (2004) stated, novices need large-scale repeated constructions of their knowledge and skills as well as scaffolding from others to construct the skills. Though the tutorial group’s students could refer back to the tutorials when they needed to, instructor support is obviously more effective.

From a different point of view, better task performance could be explained with lower state anxiety levels. In their study they conducted with college students, O’Neil, Spielberger, & Hansen (1969) found that students with high state anxiety scores made more errors on difficult materials and fewer errors on easy materials than students with low state anxiety scores. Also, according to the drive theory proposed by Spence and Taylor, “the performance of high-anxious students would be inferior to that of low-anxious students on complex or difficult tasks in which competing error tendencies were stronger than correct responses, and superior on tasks in which correct responses were dominant relative to incorrect response tendencies” (O’Neil, Spielberger, & Hansen, 1969). However, the
cognitive load levels of the groups indicated that students from both groups spent equivalent amounts of mental effort to fulfil the tasks. Therefore, we cannot rely on the principles of the drive theory to explain the difference between the task performance scores.

The redundancy effect is one of the principles of cognitive load theory. Presentation of the same information through the same channel using different formats is called the redundancy effect. The redundancy effect occurs when additional information, rather than having a positive or neutral effect, interferes with learning (Sweller, 2002). There are many different forms of redundancy. Mental activity/physical activity redundancy occurs when, for example, learning how to use a computer application by reading a text has the added physical activity of using the computer (Cerpa, Chandler, & Sweller, 1996; Chandler & Sweller, 1996; Sweller & Chandler, 1994, cited in Sweller, 2002). Either reading the text in a manual or surprisingly, physically using a computer, can be redundant and interfere with learning (Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Reder & Anderson, 1980: 1981, cited in Sweller, 2002). Based on the redundancy principle, we could expect the tutorial group students to find the tasks to be more difficult. However the two groups were found to be identical regarding cognitive load levels and students from both groups stated that they spent moderate levels of cognitive load in average during the tasks, which means cognitive overload did not occur and cognitive load did not have a significant effect on differentiation of task performances. From this finding we can infer that both tutorial and demonstration methods equally affect cognitive load at least for cases in which cognitive overload does not occur.

After analyses, the two groups were found to be equal regarding achievement test scores, recall test scores, and task durations. These results may be interpreted as proof that neither the tutorial nor the demonstration methods were superior to each other in this study in affecting achievement, recall, and task duration.

Findings regarding cognitive load levels for both groups indicate cognitive load level is not only related to task difficulty but also to other variables such as instructor support, instructional routines, and instructional method. As stated above, mental/physical activity redundancy may have affected cognitive load levels of the tutorial group to the level at which tutorials became routine for the students. On the other hand, findings on state anxiety levels can be explained with scaffolding. Unfamiliarity with the software may have required more support and yielded higher state anxiety levels in the former assignments. As students constructed new skills, they gradually became more and more skilled in fulfilling the tasks and had lower state anxiety levels.

Cognitive load levels and state anxiety levels showed moderate or weak positive correlations with each other for both groups. We can interpret this finding as students tending to feel more anxious during more difficult tasks. However, mental effort was not a strong predictor of state anxiety.

Task performance scores, however, had varying relations with cognitive load levels and state anxiety levels. For the tutorial group task performance had weak positive correlations with cognitive load and anxiety, which means cognitive load and state anxiety levels had positive weak effects on task performance. This finding may be explained with the Yerkes-Dodson Law (Smith, Sarason, & Sarason, 1982, cited in Macintyre, 1995). Parallel with this law, state anxiety level seemed to have little effect on performance and even improved it. Also a moderate level of cognitive load may have directed these students to pay more attention to the assignments, thus increasing task performance scores. In order to understand the relationship between anxiety and performance, Eysenck & Calvo (1992) proposed a theoretical model (processing efficiency theory or PET). It suggests that as state anxiety increases, worries related to evaluation and performance lower the number of cognitive resources available to complete tasks (affecting task efficiency). It is argued that worry motivates anxious individuals to allocate additional resources to tasks in order to maintain task effectiveness (Hadwin, Brogan, & Stevenson, 2005). On the other hand, task performance correlated negatively with cognitive load and had no correlation with the anxiety level of the demonstration group. We can explain this result by the students failing to remember instructions and explanations. Some of the students from the demonstration group stated they forgot the instructions while they were carrying out them. Probably students from this group forgot the harder instructions more, resulting in lower task performances.

While achievement test scores showed negative moderate correlations with cognitive load and anxiety levels for the demonstration group, recall test scores had moderate levels of positive correlations with cognitive load and state anxiety for the tutorial group. These results seem confusing but they can be explained by the difference between doing and recalling. State anxiety and increased cognitive load may have hindered students’ skills when they were in front of a computer. The same variables may have had positive effects on recall. Obviously more information is needed to arrive at firm explanation.
Correlations among task durations, cognitive load and anxiety levels were generally positive. In addition, there was a moderate level of positive correlation between task duration and state anxiety level for the demonstration group. It seems that mental effort wasted on a task had a lengthening effect on task duration. Also the state anxiety level had a lengthening effect on task duration as well. We can support this finding with the findings of Hadwin, Brogan, & Stevenson (2005), who conducted a study with nine- and ten-year-old children on the effect of state anxiety on tasks tapping the central executive, phonological, and visuo-spatial components of working memory. Although they found no differences between high and low state anxiety groups in task accuracy for any measure, children in the high state anxiety group took longer to complete the backward digit span task and reported increased mental effort in the forward digit span task, indicating some effect of anxiety on measures of performance efficiency.

Student opinions showed that tutorials provided the opportunity for individual study and they made access easier to the instructions. In general, students tended to think that tutorials decreased but not completely diminished the need for an instructor. While some students were happy with solving problems by themselves, others preferred the explanations of an instructor and the opportunity to consult the instructor in need. On the other hand, demonstration has its own advantages. Ease of comprehension of the instructions, instructor support, peer support and increased social interaction are the most significant advantages of demonstration specified by the students. In his study where he compared computer-aided and traditional instruction of computer skills, Varank (2006) too emphasizes the importance of interaction and feedback to maximize student motivation. Probably the weakest aspect of demonstration is the possibility for some students to forget the instructions.

Conclusion

In their study, Martin & Dunsworth (2007), concluded that hands-on projects, in-class activities, handouts, and PowerPoint presentations were found to be the most helpful strategies by college students. On the other hand, textbooks, discussion forms, and cooperation were found as the least helpful strategies in computer literacy instruction. Parallel to these findings, our research also showed both tutorial and demonstration methods can be used successfully with college students in computer skills instruction. Both have their own strengths and weaknesses. The most important aspect for both methods emerged as the instructor support.

I believe the conducted study has outcomes that may influence the instruction of software applications. First of all, the findings give support to the idea of instructors may use one of the two methods where appropriate. As a limitation, no answers were provided to such a question as “which method works better under which circumstances?” Instructors may use both methods in a blended way to take advantages of the two. For example, tutorials may be supported with demonstrations to increase instructor support, social interaction and decrease anxiety. Tutorials may be supportive to demonstrations to eliminate recalling difficulties which may occur during instruction. However, both methods are close to the cognitive-behaviourist approach. As the researcher I think their efficiencies are limited. Computer technology is evolving at such a rapid rate that, if an individual undertakes traditional, directive-style training in how to use a particular piece of software, that knowledge is likely to be inadequate or out-of-date in a very short period of time (Phelps, Hase, & Ellis, 2005). More constructivist methods and strategies should be employed as novice learners construct computer skills. I should mention that little literature on the topic provides mixed results. Regarding computer skills instruction, while some of the studies noted superiority of constructivist models and theories over traditional methods (e.g., Barg et al., 2000; Asan & Haliloglu, 2005), some of the studies noted superiority of traditional demonstration and practice-based methods (e.g., Akdemir & Memis, 2008). There are also studies (e.g., Park & Ertmer, 2007) that concluded no superiority of the two approaches over the other. We need to search for new ways to develop positive attitudes and self-confidence toward computers in order to maximize the effectiveness of computer skill instruction (Birol et al., 2009). I may recommend future researchers to study on integrating the two methods compared in this study with constructivist learning theory and models.

References

Adulseranee, R. (2007). The effects of using different types of multimedia presentations on Thai seventh-grade learners’ understanding of social studies text (Doctoral dissertation). Northern Illinois University, DeKalb, USA.


Prato, C. A. (2009). Biofeedback-assisted relaxation training program to decrease test anxiety in nursing students (Doctoral dissertation). University of Nevada Graduate College, Las Vegas, USA.


Student Engagement in Blended Learning Environments with Lecture-Based and Problem-Based Instructional Approaches

Ömer Delialioğlu
Department of Computer Education and Instructional Technologies, Faculty of Education, Middle East Technical University, Ankara, Turkey // omerd@metu.edu.tr

(Submitted April 5, 2011; Revised October 15, 2011; Accepted October 18, 2011)

ABSTRACT
This study investigates how blending of different instructional approaches with technology affects students’ engagement. A computer networks course was designed and implemented for the first eight weeks of the semester as a lecture-based blended learning environment and for the second eight weeks of the semester as a problem-based blended learning environment. A single group repeated measures research design was carried out to understand if there are significant differences in measures of student engagement between these two blended learning approaches. Repeated measure ANOVA analysis on the data collected from 89 students revealed that Active Learning and Total Time on Task indicators of student engagement were significantly higher in the problem-based part of the course. Interaction and Level of Academic Challenge components and course satisfaction did not show any significant differences between the two parts. Regression analysis showed that the difference in Active Learning is not due to student individual differences but rather the learning environment provided in the problem-based blended learning.

Keywords
Interactive learning environments, Teaching/learning strategies, Evaluation of CAL systems, Pedagogical issues, Improving classroom teaching

Introduction
The trend of decreasing student satisfaction from education in secondary/high school education and higher education context has drawn attention to the concept of student engagement. One of the important factors for student learning and personal development is students’ level of engagement with academically purposeful activities (Kuh, 2001). Students’ low engagement with academic activities is considered the main reason for dissatisfaction, negative experience, and dropping out of school in some of the previous research studies (Greenwood, Horton, & Utley, 2002; Legters, Balfanz, & McPartland, 2002; Perie, Moran, & Lutkus, 2005). Interventions to improve student engagement are mainly instructional solutions such as designing learning environments and utilization of engaging teaching practices.

Many personal, environmental, and instructional factors impact student achievement and personal development in educational institutions (Burkman, Tate, Snyder, & Beditz, 1981; Everson & Millsap, 2004; Walberg, 1984; Wenglinsky, 2002). Among them, educators have control over only instructional practices. By designing and implementing various instructional environments and practices, student learning and development could be improved. Research in higher education and secondary/high school education context alike has agreed that students’ engagement with academically purposeful activities is one of the important factors for student learning and personal development in traditional and technology enhanced learning environments (Astin, 1993; Pace, 1980; Pascarella & Terenzini, 1991; Jelfs, Nathan, & Barrett, 2004; Ginns & Ellis, 2007). Therefore, it is highly recommended that educational institutions and instructors direct their energy and resources to the methodologies and technologies to improve student engagement in their institutions (Hu & Kuh, 2002).

While secondary/high school education literature defines student engagement in three categories, higher education literature provides an umbrella definition for student engagement as students’ involvement with academically meaningful activities (Kuh, 2001). Regardless of the definition and the context, the most important question to answer is, “What keeps students engaged in schools and colleges?” Reviewing voluminous research in higher education context, Chickering and Gamson (1987) put forward a framework to ensure students’ engagement; “Seven Principles for Good Practice in Undergraduate Education”. According to this framework, students are more engaged when the instruction (a) increases the contact between student and faculty, (b) provides opportunities for students to work in cooperation, (c) encourages students to use active learning strategies, (d) provides timely feedback on
students’ academic progression, (e) requires students to spend quality time on academic tasks, (f) establishes high standards for acceptable academic work, and (g) addresses different learner needs in the teaching process. Similarly, Brophy and Good (1986) provided a review of research on teaching and teaching practices that aim to improve student learning and achievement. Although Brophy and Good (1986) did not use the term engagement, they provided recommendations similar to Chickering and Gamson (1987) for providing engaging instruction for students. According to Brophy and Good (1986), teaching practices that encourage active learning strategies, establish collaborative student work, contain challenging tasks, and provide prompt feedback help improve student achievement and learning in schools.

Past findings from various educational researches confirm that student engagement is an important construct for learning and personal development (Astin, 1993; Pace, 1980; Pascarella & Terenzini, 1991; Jelfs, Nathan, & Barrett, 2004; Ginnis & Ellis, 2007). Designing such learning environments requires utilization of instructional design strategies that address principles of student engagement. The combination of face-to-face and online learning environments gave rise to a new learning environment called blended learning environments. Combining face-to-face and online learning environments has potential to provide a learning environment where student engagement opportunities are more than using only one type of learning environment (Boyle, Bradley, Chalk, Jones, & Pickard, 2003; Brennan, 2003; Osguthorpe & Graham, 2003).

Blended learning was given different names throughout the years, such as hybrid instruction, mediated learning, technology enhanced instruction, web enhanced instruction, and web assisted instruction. Currently, blended learning seems to be the “de facto” term to refer to these mixed modes of learning environments. On the other hand, the term blended learning is criticized because the blend occurs not in the learning but in the teaching. Oliver and Trigwell (2005) proposed that the term has to be “blended pedagogies”, “blended teaching” or “learning with blended pedagogies”. There are different definitions of blended learning in the literature. Blended learning environment are defined as a combination of face to face and online learning environments to utilize strengths of both (Delialioglu & Yildirim, 2007; Osguthorpe & Graham, 2003). The definition of Garrison and Kanuka (2004) focus on the way of “blending” the learning experiences. They define blended learning as “the thoughtful fusion of face-to-face and online learning experiences”.

Bliuc, Goodyear, and Ellis, (2007) reviewed research studies done on blended learning in higher education. They classified research done on blended learning environments throughout the past years according to the methodological choices they have used to investigate students’ experiences. Accordingly, most of the research on blended learning was classified as case-studies, survey based studies and comparative studies. There was a small group of research studies using a holistic approach. Having reviewed voluminous research, Bliuc, Goodyear, and Ellis (2007) argued that educationally useful research on blended learning have to investigate the relationship between different modes of learning and how they could be integrated. As it is the case in the current study, they expect research on blended learning to yield usable evidence on good quality student learning experiences.

There are a few studies that have focused on blending technology with problem-based learning instructional approach rather traditional face-to-face lecture based instruction. Donnelly (2006) performed a case study focusing on using problem-based blended learning, which was called “blended problem-based learning” in the study, as the dominant pedagogical approach to integrate face-to-face classroom instruction to e-learning environment. The study reflected the learning experiences of academic staff (lecturers, librarians, and learning technology support staff) in a problem-based learning context supported with online learning tools and technologies. The research suggested that the principles of constructivism and engagement are critical factors to consider when designing problem-based blended learning environments. Donnelly (2010) conducted a qualitative study in higher education context on how to combine instructional strategies in face-to-face and computer-mediated environments so as to use their strengths but avoid their weaknesses. The research was specifically interested in understanding the role of technology as a tool to support interaction of learners with each other and with the content in a problem-based learning context. The findings of the study pointed that combination of technology support in the form of interactive media and problem-based learning could be seen as a formidable combination but they are complementary rather than collide.

There are cases in educational research literature where problem-based learning with technology support was used to improve the quality of student experiences in learning (Keppell, 2005; Uden & Beaumont, 2006). These cases were not specifically called problem-based blended learning environments, but the underlying paradigm of technology integration to instruction is the same as blended learning courses.
The purpose and questions of the study

Blended learning environments have been utilized in higher education context for more than ten years, but there are few studies that examine student engagement in these environments. The literature has a gap in terms of the effect of different blending approaches on student engagement (Keppell, 2005; Uden & Beaumont, 2006; Donelly, 2006; Donelly, 2010). Therefore, there is a need to extend the line of inquiry of student engagement and approaches for blending technology with different instructional approaches. The purpose of this study is to compare student engagement within a lecture-based blended learning environment to that within a problem-based blended learning environment. The research questions that guide this study are:

1. Are there significant differences in the indicators of student engagement measured as Active Learning, Student Interaction, Level of Academic Challenge and Time on Task between a lecture-based blended learning environment and a problem-based blended learning environment?
2. Is there a significant difference in students’ course satisfaction between a lecture-based blended learning environment and a problem-based blended learning environment?
3. Do students’ individual differences (student demographics attributes, student motivation and ability attributes) impact student engagement in a lecture-based blended learning environment and a problem-based blended learning environment?

Method

The study used a repeated measure research design to compare student engagement. A third year computer networking course in a pre-service teacher education program was implemented as a blended learning course in spring 2008 and spring 2009 semesters. An online learning environment was blended with a lecture-based instruction for the first eight weeks of the semester and then with a problem-based instruction approach for the second eight weeks of the semester. Students took two surveys at three different times in two consequent years during the study. First, an entry survey was administered to measure student abilities and motivational aspects at the beginning of the study. Second, an engagement survey was conveyed at the end of the first half of the course to measure student engagement in blended learning with problem-based instruction. Finally, the same engagement survey was completed at the end of the blended learning with problem-based instruction part of the class. The same design was used for both years as summarized in Table 1.

<table>
<thead>
<tr>
<th>Beginning of the Semester</th>
<th>Treatment 1 (First 8 Weeks)</th>
<th>Mid-Semester Measure 1</th>
<th>Treatment 2 (Second 8 Weeks)</th>
<th>End of Semester Measure 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry Survey</td>
<td>Lecture-based blended learning environment</td>
<td>Engagement Survey</td>
<td>Problem-based blended learning environment</td>
<td>Engagement Survey</td>
</tr>
<tr>
<td>(Demographics, Motivation &amp; Ability factors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There were two types of variables in the study: (i) dependent variables, (ii) independent variables. The variables and measures of the study are listed in Table 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Students’ Engagement;</td>
<td>Engagement Survey</td>
</tr>
<tr>
<td>Active Learning</td>
<td>Interval Scale</td>
</tr>
<tr>
<td>Student Interaction</td>
<td>Interval Scale</td>
</tr>
<tr>
<td>Level of Academic Challenge</td>
<td>Interval Scale</td>
</tr>
<tr>
<td>Time on Task</td>
<td>Interval Scale</td>
</tr>
<tr>
<td>Course Satisfaction</td>
<td></td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Treatment 1: Lecture-based blended learning environment</td>
<td></td>
</tr>
<tr>
<td>Treatment 2: Problem-based blended learning environment</td>
<td></td>
</tr>
<tr>
<td>Student Individual Factors</td>
<td>Entry Survey</td>
</tr>
</tbody>
</table>
Participants

Participants of the study were 93 junior pre-service computer teachers. A group of 38 students enrolled in the course for the spring 2008 semester, and a group of 55 enrolled in the course for the spring 2009 semester. Of these participants, 63 were male and 30 were female. Students were invited to take the surveys through e-mails. The first two surveys were taken by all 93 students. The response to the third survey was lower – a total of 89 students took all three surveys. To provide proof of equivalent for the two consecutive year’s students to be used as a single group a t-test was run on motivation, GPA, and technical skills student individual factors between the years. The results of the t-test are provided in table 3. The results of the analysis indicate no significant difference between the years.

<table>
<thead>
<tr>
<th>F</th>
<th>df</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>0.17</td>
<td>87</td>
</tr>
<tr>
<td>GPA</td>
<td>0.72</td>
<td>87</td>
</tr>
<tr>
<td>Technical Skills</td>
<td>0.32</td>
<td>87</td>
</tr>
</tbody>
</table>

The results of the one-way ANOVZ indicated no significant difference between years. For further analysis in finding answers for the research questions, a repeated measure ANOVA was used on the collected data. Year and treatment were tested for between-subjects effect.

Context

The study was carried out at a large public university. The research context was a junior level computer networking course offered to pre-service computer teachers. Although the course was a part of the curriculum and a mandatory course, participation in the study was voluntary. The course was designed as a blended learning environment where students and instructor meet weekly face-to-face and received the course materials online. The instructional materials were delivered to instructors and students using a propriety online course management system via the Internet. There were differences between the blended environments in terms of the teaching-learning tools utilized. The differences are summarized in Table 4.

<table>
<thead>
<tr>
<th>Instruct. Technologies &amp; Learning Tools</th>
<th>Lecture-Based Blended Learning Environment</th>
<th>Problem-Based Blended Learning Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Content</td>
<td>Students had to read weekly before attending the class.</td>
<td>What and when to read was not specified.</td>
</tr>
<tr>
<td>Multimedia Components</td>
<td>Animations and simulations were integral to the online content.</td>
<td>Were integral to the online content but students were not required to do them.</td>
</tr>
<tr>
<td>Online Assessment</td>
<td>Students had to take weekly online chapter quizzes.</td>
<td>Not required</td>
</tr>
<tr>
<td>Lab Activities</td>
<td>Students were required to do the hands-on activities as in their lab-sheets. Activities included procedural information on what device to use and how to configure.</td>
<td>No lab-sheets</td>
</tr>
<tr>
<td>Network Simulation Software</td>
<td>Students had to use it for the lab activities.</td>
<td>Is the main working area of students to find the problems and provide</td>
</tr>
</tbody>
</table>
In blending the online content with either lecture-based or problem-based instructions, there were face-to-face classroom meetings, but the classroom activities were different. The course met twice a week. In blending the lecture-based approach with online content, the activities were as follows: (i) the instructor presented the content, (ii) discussion of the content took place among the instructor and other students, and (iii) hands-on laboratory activities were conducted. In blending of problem-based approach, the classroom activities were as follows: (i) an ill-structured case with problems related to the week’s content was provided and explained, (ii) students were required to work in groups of two to discuss the case and determine what they needed to know, and (iii) students were asked to write a problem statement and possible solutions. Students were required to deliver a report with possible solutions in two weeks.

**Procedures**

Basic concepts of networking were covered throughout the whole blended course. The topics that were covered in the course are listed in Table 5.

<table>
<thead>
<tr>
<th>Table 5. Topics covered in the blended course</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lecture-Based Blended Learning Environment</strong> (First 8 weeks)</td>
</tr>
<tr>
<td>Living in a Network Centric World</td>
</tr>
<tr>
<td>Communicating over the Network</td>
</tr>
<tr>
<td>OSI Transport Layer</td>
</tr>
<tr>
<td>TCP/IP Protocol Suite and IP Addressing</td>
</tr>
<tr>
<td>Addressing the Network - IPv4</td>
</tr>
<tr>
<td>Ethernet Fundamentals</td>
</tr>
</tbody>
</table>

During the lecture-based blended learning, the instructor utilized teacher centered methods such as presentation of information, demonstration, and chapter quizzes. During the problem-based blended learning, the students were given ill-structured cases with problems, were required to solve the problems of the case, and provide a report. The report should include information they needed, what activities they did, what they have learned and solution(s) to the cases.

During the first eight weeks of the semester the instructor elaborated online readings with presentations on the important concepts. Students were encouraged to participate in the lectures through questions and answers. Students were required to attend laboratory hours and complete hands-on activities.

The second part of the course was designed as a blend of problem-based approach and online content. In the classroom and the computer laboratory, students were given an ill structured case created in a network simulation software environment. First, the case and the problem were explained by the instructor in the classroom. Then students were required to work in groups of two. Students continued to work on the case during laboratory hours. They could ask questions to their lab assistants about the requirements and the concepts but not on the solution of the problem. Parallel to the structure of the online content, a total of 5 cases were given throughout the 8 weeks of the problem-based blended learning environment. A sample of an ill-structured case requiring the students to learn IP concepts and IP addressing is provided in Figure 1.
Data collection instruments & measures

Two survey instruments were used to collect data during the course periods, the Entry Survey and the Student Engagement Survey. Both surveys were administered online to the students. Since students took the surveys at three times during the semester, it was important to track them throughout the semester. Before taking each survey, students entered their student ID numbers. Student ID numbers were deleted after merging the data from all three surveys so that no student could be identified in the survey.

Entry survey

At the beginning of each semester the course was offered, students took an Entry Survey to collect information about their demographics, motivation, and ability factors. Students were invited to take the online survey through e-mail and through their instructor’s encouragement in the face-to-face session of the class. A second reminder e-mail was sent to the students to take the survey. The survey was open to students for a week. A total of 38 and 55 students took the Entry Survey during the 2008 and 2009 spring semesters respectively.

Reliability and validity measures of the Entry Survey were established in a large scale evaluation study using the same scales (Dennis et al., 2006). The Entry Survey measures were composed of three categories: students’ demographics, motivation, and ability factors. Demographics measures consist of student gender, parents’ education level and intended educational degree categorized as binary variables. Student motivation was measured using Eccless and Wigfield’s (2002) 7-item expectancy and value scale. The response scale for the items was a 5-point Likert scale. In order to check the reliability and the internal consistency of the survey Cronbach alpha value was obtained as 0.83 for the motivation scale. Student ability was measured in two ways; first, self-reported student GPAs and second, student technical abilities. Students were simply asked to report their cumulative grade point average on a scale of 4.00. Students’ technical ability was measured using 5 items on a 7-point Likert scale. The
students were asked how many times they performed computer and networking related activities in a year. The Cronbach’s alpha score was 0.86 for the technical ability scale.

**Student engagement survey**

Students took a second survey that measured their level of academic engagement with the instructional activities at two different times; after completing the lecture-based blended learning, and after the problem-based blended learning. The same method as in the Entry Survey was used to invite students to take the survey; that is students received an invitation e-mail with a link to the survey. The survey was open for a week to students’ access. At the end of the problem-based part of the course students took the same Student Engagement survey. The same procedure was followed to invite and encourage students to take the survey. A total of 89 students took the second engagement survey during the two semesters the course was offered.

The level of student engagement was measured using a survey developed by the National Survey of Student Engagement (NSSE). The Student Engagement Survey is a tool to measure the level of student engagement with five benchmarks: (i) level of academic challenge, (ii) student-faculty interaction, (iii) active and collaborative learning, (iv) personal development and (v) supportive learning environment (Kuh, 2001). In this study, student engagement was measured with three constructs of the survey; Active and Collaborative Learning, Student-Faculty Interaction, and Level of Academic Challenge. Two factors were added to the Student Engagement Survey; Total Time on Task and Course Satisfaction. Total Time on Task is the sum of time in hours the students spend doing specific learning activities. The Course Satisfaction asks questions on the quality of the instruction and overall learning environment. The items in the Active and Collaborative Learning construct ask students about frequency of learning strategies they use in and out of school during the course period via six items with a 5-point Likert scale. The Student-Faculty Interaction construct measured the level of interaction between the student and faculty via five items with a 5-point Likert scale. Interaction between students and faculty is another contributor towards student learning and personal development; therefore, it is expected that students with more interaction scores have higher engagement with the academic activities. Level of Academic Challenge refers to how hard students worked to meet their instructor’s expectations in the course. Level of Academic Challenge was measured with a five-point Likert scale item. The alpha reliability scores of Level of Academic Challenge, Active Learning and Interaction scales were found to be 0.83, 0.81 and 0.84 respectively.

**Data analysis**

The main methodology of the study was the single group repeated measures design. Two groups of students in two consecutive years received the same two treatments and their level of engagement was measured with the student engagement survey at the end of each treatment. In comparison to the conventional repeated measures design, the single group repeated measures design offers several advantages. First, if only a small number of participants are available in a study, the single factor design could be used. Second, the same participants receive two different treatments; therefore, no group differences exist due to the entry conditions in the comparison and in the experiment groups, which reduces the impact of unchecked confounding factors between the groups (Gliner, Morgan, & Harmon, 2002). In order to understand the differences between the two different blended learning environments with respect to student engagement and to control the effect of year on the measures students’ scores in the Student Engagement Survey were compared with a repeated measure ANOVA where year and treatment were used as between groups’ factors. Components of engagement which were found to be significantly different between the treatments were further tested for the impact of student individual factors to engagement indicators through regression analyses. Since gender, parents’ education level, intended final educational degree were binary variables, they were entered to the regression model as dummy variables as required in such analysis.

**Results**

Among the enrolled 93 students, a total of 89 students took both engagement surveys. As Table 6 indicates, students who participated in the study showed a balanced distribution. While 63 of the students were male, 30 of them were female, which can be considered as a high male to female ratio in a classroom with heavily technical content.
Educational level of the students’ parents showed an interesting picture; a significant number of the students indicated that their mother or father did not have high school degree. Finally, almost half of the students intended to have bachelor’s degree as their final educational degree; the other half aimed to have graduate degrees.

Table 6. Distribution of student demographics attributes

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>63</td>
<td>67.8%</td>
</tr>
<tr>
<td>Female</td>
<td>30</td>
<td>32.2%</td>
</tr>
<tr>
<td>Mother’s Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No High School</td>
<td>64</td>
<td>68.8%</td>
</tr>
<tr>
<td>High School or More</td>
<td>29</td>
<td>31.2%</td>
</tr>
<tr>
<td>Father’s Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No High School</td>
<td>49</td>
<td>52.7%</td>
</tr>
<tr>
<td>High School or More</td>
<td>44</td>
<td>47.3%</td>
</tr>
<tr>
<td>Intended Final Educational Degree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>48</td>
<td>51.6%</td>
</tr>
<tr>
<td>Graduate</td>
<td>45</td>
<td>48.4%</td>
</tr>
</tbody>
</table>

Student motivation was measured with Eccles and Wigfield’s (2002) expectancy-value scale at the beginning of the semester in the Entry Survey. Student motivation and ability factors measured on scale value are presented in Table 7. The mean motivation score for the participating students indicated that students were highly motivated at the beginning of the class. They expected to learn a lot and they valued the course content for their education and career life. Two ability factors, GPA and technical skills were also measured at the beginning of the semester. The mean score for GPA for all participants was 2.74 over 4.00, which indicates an average achievement level for the class. Student technical skills for basic computer and networking tasks were measured on a five-item seven-point Likert scale. The mean score for the technical skills construct was 2.58 with 1.42 standard deviation, which indicates a below average score. Considering that this information was collected at the beginning of the semester, participating students were average students with high motivation to learn.

Table 7. Distribution of student motivation and ability attributes

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>89</td>
<td>2.85</td>
<td>5.00</td>
<td>4.21</td>
<td>0.51</td>
</tr>
<tr>
<td>GPA</td>
<td>89</td>
<td>1.79</td>
<td>3.85</td>
<td>2.74</td>
<td>0.48</td>
</tr>
<tr>
<td>Technical Skills</td>
<td>89</td>
<td>0.61</td>
<td>6.00</td>
<td>2.58</td>
<td>1.42</td>
</tr>
</tbody>
</table>

The mean scores for engagement indicators during the course are provided in Table 8. The mean score and standard deviation for student engagement outcomes are showed as Mean 1 and Standard Deviation 1, which were measured after the lecture-based blended learning part of the course. Mean 2 and Standard Deviation 2 were the mean score and standard deviation for student engagement outcomes measured after the problem-based blended learning part of the course. Total time on task was measured with hours and all other outcomes were measured on a 5-point Likert scale. Considering that the maximum value on the scale was 5 (strongly agree), Active Learning and Interaction outcomes of student engagement remained relatively low in both types of learning environments in the course.

Table 8. Descriptive statistics for engagement outcomes of the blended courses

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1 Lecture-Based</th>
<th>Treatment 2 Problem-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean 1</td>
</tr>
<tr>
<td>Active Learning</td>
<td>89</td>
<td>2.29</td>
</tr>
<tr>
<td>Interaction</td>
<td>89</td>
<td>2.37</td>
</tr>
<tr>
<td>Level of Academic Challenge</td>
<td>89</td>
<td>3.24</td>
</tr>
<tr>
<td>Total Time on Task</td>
<td>89</td>
<td>14.24</td>
</tr>
<tr>
<td>Course Satisfaction</td>
<td>89</td>
<td>3.93</td>
</tr>
</tbody>
</table>

As Table 8 shows, level of student engagement outcomes and satisfaction with the course were different in the lecture-based and problem-based portions of the blended course. The mean scores of engagement outcomes and total time on task were higher in the latter. In order to understand whether these differences were significant, repeated
measure ANAVO were performed for each of the outcome measures of the student engagement. Results of analysis were presented in Table 9.

As shown in Table 9, significant differences existed in some of the student engagement measures between lecture-based and problem-based blended learning parts of the course. Analysis revealed that there was no statistically significant difference between groups in terms of year. In terms of treatment, Active Learning and Total Time on Task were significantly higher in the latter, whereas Interaction, Level of Academic Challenge, and course satisfaction did not show any significant differences between the two.

### Table 9. Results of repeated measure ANAVO analysis for engagement outcome pairs

<table>
<thead>
<tr>
<th>Source</th>
<th>F</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active learning</td>
<td>21.85</td>
<td>1</td>
<td>0.000a</td>
</tr>
<tr>
<td>Interaction</td>
<td>2.16</td>
<td>1</td>
<td>0.429</td>
</tr>
<tr>
<td>Level of Academic Challenge</td>
<td>1.90</td>
<td>1</td>
<td>0.403</td>
</tr>
<tr>
<td>Total Time on Task</td>
<td>65.00</td>
<td>1</td>
<td>0.000a</td>
</tr>
<tr>
<td>Course Satisfaction</td>
<td>1.34</td>
<td>1</td>
<td>0.170</td>
</tr>
<tr>
<td><strong>Year</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active learning</td>
<td>0.34</td>
<td>1</td>
<td>0.854</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.03</td>
<td>1</td>
<td>0.956</td>
</tr>
<tr>
<td>Level of Academic Challenge</td>
<td>0.13</td>
<td>1</td>
<td>0.910</td>
</tr>
<tr>
<td>Total Time on Task</td>
<td>0.10</td>
<td>1</td>
<td>0.753</td>
</tr>
<tr>
<td>Course Satisfaction</td>
<td>0.05</td>
<td>1</td>
<td>0.822</td>
</tr>
</tbody>
</table>

*a p<0.01

Active Learning refers to students’ actively searching for information to make sense of the content that is provided in a course. Traditionally, the lecture-based method of instruction gives students a passive role in their learning process. Whereas, the problem-based learning method of instruction encourages students to actively engage in information seeking and in meaning making process. Students reported significantly higher use of active learning strategies during the problem-based blended learning part of the course in comparison to the lecture-based blended learning part of the course. Problem-based learning increased student engagement in blended learning environments similar to traditional learning environments.

Another significant finding was the difference between the total time spent on academic activities during the lecture-based and problem-based parts of the blended course. Students reported significantly higher time spent during the problem-based blended learning part of the course. In comparison to the traditional lecture-based methods of instruction, problem-based learning activities required students to actively engage in the learning process. Typical learning activities include but not limited to information seeking, collaboration with other students, and synthesizing the information from various resources. Therefore, similar to the previous research findings in traditional learning environments, in blended learning environments, problem-based learning methods engaged students with academic activities better than lecture-based methods do.

Other indicators of student engagement, Interaction and Level of Academic Challenge, do not differ significantly between the two methods of instruction. Since the problem-based learning methods inherently encourage students to interact, Student Interaction was expected to be higher in the problem-based part of the course in comparison to the lecture-based part. Although the mean score for student interaction was higher in the problem-based blended learning part than the one in the lecture-based blended learning part, the difference was not significant.

Similarly, Level of Academic Challenge did not show any significant difference between the two methods of instruction. Although the students indicated that they experienced a greater level of academic challenge in the problem-based part of the course than in the lecture-based part, it was not statistically significant.

Students were highly satisfied with both parts of the course and no significant difference exists between the two. Student satisfaction was high for both types of the blended learning environments.

Comparing student engagement scores between lecture-based and problem-based blended learning environments resulted in significant differences in Active Learning measure of student engagement and students’ time spent with academic activities. The other measures of student engagement did not yield significant differences between two...
methods of instruction. Time spent in academic activities in lecture based and problem based methods were expected to be significantly different in favor of problem based learning method, and the results in Table 9 confirms this anticipation.

Having found significant difference of student engagement between lecture-based and problem-based blended learning methods, the next step of the analysis was to understand what individual factors were important on student engagement, measured as Active Learning in both methods of instruction, which answers Research Question 3.

A model was established to explore the impact of student individual factors on Active Learning. Two regression analyses were run using the same model. In the regression model, while dependent variable was active learning measure of student engagement, independent variables were student individual factors measured in the Entry survey presented in Table 6 and Table 7. The binary variables entered the regression analysis as dummy variables as required in such analysis.

Table 10 shows regression analysis results for Active Learning measure in lecture based method. Other than gender, none of the student individual factors have significant impact on student active learning in the lecture-based blended learning environment. According to the results, male students benefit more than female students in terms of active learning activities in lecture-based blended learning environment.

Table 11 shows regression analysis results for active learning measure in problem-based blended learning environment. Similar to the results of the lecture-based method, none of the student individual factors impact student engagement.

Except for gender factor, in both analyses, student individual factors show no significant impact on student engagement in both learning environments. These results emphasize that regardless of student individual difference included in the regression model, problem-based blended learning environment engaged students with activities better than lecture-based blended learning environment. Additionally, lecture-based blended learning environment favors male students to engage in active learning, requiring instructors to develop activities to improve female students’ engagement with academically meaningful activities.

Conclusion and discussion

The purpose of this study was to compare in a blended learning environment context the level of student academic engagement through lecture-based with the level of engagement through problem-based learning strategies utilized.
No significant differences for Level of Academic Challenge and Course Satisfaction could be explained with the use of online materials in a blended learning environment. Previous research showed that instructional materials used in the current study were already difficult and challenging for students (Bichelmeyer et al., 2006). As measured at the beginning of the course, participating students have below average technical skills. Combination of a demanding technical course and students with low level technical skills results in a similar level of academic challenge in both types of instructional methods applied to the blended learning. Regardless of the utilized instructional strategy, students were highly satisfied with the course configured as a blended learning environment. One explanation for this outcome could be that the course materials include various resources such as multimedia, simulations, hands-on activities, and games delivered over the Internet with live classroom sessions. An advantage of the blended learning environment is that, while students had the freedom to study at their own pace, they also had the opportunity to discuss the online material and laboratory activities with their peers and the instructor to perform the hands-on activities.

In comparison to the lecture-based part of the course, students were more engaged with active learning strategies and they spent more time on academic activities during the problem-based learning part of the course. On the other hand, the student interaction and the level of academic challenge perceived were similar in both instructional strategies. Additionally, students were equally satisfied with the course regardless of the instructional strategy used.

Problem-based learning as an instructional strategy in this study’s blended context was effective to improve student engagement, as well. However, interestingly, students’ level of interaction and level of academic challenge were very similar in both types of instructional strategy in the blended course format. In the traditional learning environments, the problem-based learning method increases the level of interactions between students and the instructor due to the tasks during the learning process (Hmelo-Silver, 2004; Prince, Eijs, Boshuizen, Vleuten, & Scherpbier, 2005). The use of blended learning environment has the potential to help close this difference between the problem-based and the lecture-based courses.

Active Learning is significantly higher in problem-based blended learning environment in comparison to lecture-based blended learning environment regardless of the student individual differences. This result heralds good news for instructors who plan to utilize problem-based blended learning method in their courses. First, problem-based blended learning environment has more power to engage students with meaningful learning activities; therefore using this learning approach will increase students’ engagement with meaningful academic activities. Second, regardless of students’ individual factors, all students engage with academically meaningful activities equally well in the problem-based environment.

Blended learning environments hold promise for the student learning and instructor practices. Future research needs to investigate instructor practices in blended learning environments and their impact on student engagement with large scale studies. Additionally, studies are needed to understand how blended learning environments impact teachers’ daily tasks and teaching practices.

As in any social research, this research is also bound with some limitations. First, the study employs the single group repeated measures design; therefore, it uses one group of subjects and they receive two different treatments consecutively. A disadvantage of this method is that subjects may have the effect from the previous treatment. Second, the dependency of the findings on self-report data and lack of controlling for students’ learning outcomes could have limited the results of the study. Therefore any generalization from this study should be carefully reviewed.

References


Development Patterns of Scientific Communities in Technology Enhanced Learning

Manh Cuong Pham, Michael Derntl and Ralf Klamma

RWTH Aachen University, Advanced Community Information Systems (ACIS), Aachen, Germany // pham@dbis.rwth-aachen.de // derntl@dbis.rwth-aachen.de // klamma@dbis.rwth-aachen.de

ABSTRACT
Conferences play an important role in scientific community building, and in cooperation and dissemination related to Technology Enhanced Learning (TEL). In this article we analyze the communities of five major conferences in TEL, including ICALT, ECTEL, ICWL, ITS, and AIED, with the aim of understanding the development patterns of these TEL communities. This is achieved through social network analysis and time series analysis applied to the co-authorship and citation networks of these conferences. In addition, we compare the development pattern of TEL conference communities to benchmarks drawn from long-established and highly successful conferences on database research. The analyses of the social network parameters of the conference communities generated several insights. We found that TEL conferences exhibit a mixed development pattern of young, emerging conferences that are still in the process of developing their communities. We also found that the more interdisciplinary conferences in our data set exhibited a slower rate of community development compared to those conferences with more focused topics. Regarding the practical implications of these findings, we have offered some recommendations to different stakeholders including conference chairs and key authors.

Keywords
Technology enhanced learning, Conferences, Community development, Social network analysis, Time series

Introduction
Technology Enhanced Learning (TEL) is an emerging research area in computer science. Similar to other sub-disciplines of computer science, conferences have a dominant role to play in the communication of TEL research. According to Microsoft Academic Search, there are 58 conferences in contrast to 18 journals in the computer education category as of April, 2012. Such domination raises questions regarding an understanding of the communities of conferences and their development patterns, in order to have an overview of the current research work in the TEL area. For researchers, understanding the community means getting to know the research environment, which leads to self-adaptation and, hopefully, improvement in the field. For conference organizers and stakeholders, an overview of their communities is important for maintaining, cultivating and promoting conferences. The application of Social Network Analysis (SNA) in the field of digital library research is very promising in terms of knowledge discovery (Pham & Klamma, 2010; Pham, Klamma & Jarke, 2011). In particular, the structure of scientific collaboration can be researched in great detail using SNA associated with two distinct data sets: the co-authorship graph and the citation graph. The co-authorship graph reveals the contribution structures of a scientific community by disclosing who has collaborated with whom in terms of co-authoring papers. The citation graph discloses the influencing areas, conferences, and journals of a conference in terms of cited papers. Together, the two graphs allow a detailed analysis of the knowledge structure and flows within a particular scientific community, but also an analysis of the knowledge flows between adjacent scientific communities.

SNA has been proposed as a means of studying TEL communities. Kienle and Wessner (2005; 2006) as well as Hoadley (2005), studied the Computer Support for Collaborative Learning (CSCL) community using statistical analysis and the visualization of citation and collaboration data of several CSCL conferences, including also the program committee members and geographical data. They analyzed the development of the community and identified key members using simple statistics such as the number of participants over time, the proportion of new and old members, as well as the geographical distribution of the members. Similarly, Ochoa, Mendez and Duval (2009) analyzed the ED-MEDIA (World Conference on Educational Media and Technology) community. Besides statistical measures, they also used SNA metrics such as betweenness centrality to rank authors. Another work on the TEL community specific to a single conference series was the analysis of EC-TEL (European Conference on Technology Enhanced Learning) by Reinhardt, Meier, Drachsler and Sloep (2011).
SNA has also been applied in knowledge mapping research within scientometrics in order to understand the organization of scientific knowledge of all sciences as well as of a single discipline. Morris and McCain (1998) explored the interdisciplinary nature of medical informatics and its internal structure using inter-citation and co-citation analysis. A combination of Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) data was used in this study. McCain (1998) performed the co-citation analysis for journals in the field of neural network research. Cluster analysis, principal component analysis and multidimensional scaling (MDS) maps were used to identify the main research areas. Regarding the computer science discipline, Ding, Chowdhury and Foo (2000) studied the relationship between journals in the area of information retrieval using the same techniques. Based on the SciSearch database, Tsay, Xu and Wu (2003) mapped semiconductor literature using co-citation analysis. The data sets used in these studies were rather small, ranging from tens to several hundred journals. In more recent work, Boyack, Börner and Klavans (2007) mapped the structure and evolution of chemistry research over a 30-year time frame. Based on a general map generated from the combined Science Citation Index Expanded (SCIE) and SSCI from 2002, the authors assigned journals to clusters using inter-citation counts. Journals were assigned to the chemistry domains using Journal Citation Ranking (JCR) categories. Then, maps of chemistry research during different time periods and at different domain levels were generated to expose the changes that have taken place over the 30 years of the development of chemistry research.

In this article, we are concerned with a different aspect of communities of TEL conferences, that is the dynamics and patterns of community development. Unlike journals, conferences facilitate communication between participants through face-to-face meetings, academic presentations and dissemination activities. A systematic comparison of the key features of scientific community shows that, depending on the duration of existence, different conferences exhibit different development patterns. Therefore, communication patterns and community building processes are of special interest. Through an analysis of community development patterns, the main purpose of this paper is to raise the awareness of conference organizers and stakeholders with regard to these development dynamics. Consequently, another aim is to provide information that facilitates the identification of strong and weak indicators within their community, and to provide hints for improvement.

Background and methodology

Conference community development model

We have proposed a model to explain the development pattern, as well as the collaborative and citation behavior in conferences and journals in Pham, Klamma and Jarke (2011); the model is displayed in Figure 1. We have shown in this previous study that this model can be used to describe and explain the community building process of many conferences in different areas of computer science in terms of co-authorship and citation networks.

The co-authorship network of a conference series consists of authors as nodes. There is an edge between two authors if they have co-authored at least one paper published in a conference event in that series (note that we use “event” to refer to one specific conference in a conference series, e.g., the ICALT 2011 event in the ICALT conference series). In the co-authorship network there are initially few connections between authors (born phase). After some events, author groups become apparent in the network (bonding phase), which are — in the best case — gradually integrated through publications that involve authors from more than one group (emergence phase). Over time, a conference series then typically forms a network topology that features a strongly connected core group of authors that is connected to other smaller groups (focused topology). Alternatively, the co-authorship can develop into an interdisciplinary topology where several groups are connected via some gatekeepers, but where there is no core group. It can also develop toward a hierarchical topology which exposes some “super gatekeepers” who connect a hierarchy of groups.

The citation network of a conference series is a graph whose nodes represent papers which were cited by papers published by that conference series. A directed edge from a citing paper to a cited paper represents a citation. Essentially, the citation network represents the body of literature cited by papers of a particular conference series. For the citation network, the development process may be different. When the conference series is focused at the beginning, its papers tend to cite a body of fundamental literature in the field, and the citation network should be very dense and have a large connected component (focused topology). When the conference series has an interdisciplinary nature, the citation network will contain several components and the connections between these
components may either be found or they may not. If the core topic of the series is still developing, the citation network may remain in the bonding or emergence phases.

As an instantiation of the development model, Figure 2 illustrates the development of the ICALT co-authorship network since its first event in 2001, in three-year intervals. Each snapshot shows a network composed of authors (nodes) and their accumulated co-authorship relationships (edges). At the inaugural event in 2001 (born phase; note that ICALT actually started off as the International Workshop on Advanced Learning Technologies, IWALT, at Massey University in New Zealand in the year 2000; however the first DBLP-indexed conference event in the ICALT series took place in 2001) we see many “isolated” nodes in the network representing authors who had one or more papers published without any co-authors. Small groups are formed by authors who have co-authored one or more papers with at least one co-author. Three years later, at ICALT 2004, we witness the existence of some larger co-author groups, and also some bonding among author groups. In 2007, some larger author groups are already clearly discernible, and the network also starts to form a clearly visible large component of core authors in the emergence phase. By 2010, the largest component is beginning to actually deserve the label “giant component” and we see that many members of the giant component have co-authorship ties to other authors and author groups on the periphery of the giant component. Although it is evident at the bottom of each network snapshot that the network includes a large pool of unconnected authors and author groups, the ICALT author network tends toward developing a focused topology.

![Figure 1. Development model for conference’s community](image1)

![Figure 2. Illustration of the community development model by example of ICALT](image2)
Time series analysis methodology

To qualitatively characterize the development process of the TEL community in relation to this development model, we applied a time series analysis on the networks to reveal the following social-network parameters over time: densification law (Leskovec, Kleinberg & Faloutsos, 2007), clustering coefficient, maximum betweenness, largest connected component, diameter, and average path length (Wasserman & Faust, 1995). These parameters, which are explained in the following paragraphs, enable us to explain the community building process in Figure 1. To interpret the shape of the community, one needs to use a combination of all of these parameters. Formally, given the network $G = (V, E)$, where $V$ is the set of vertices or nodes, and $E$ is the set of edges, the above network metrics are defined as follows:

**Densification law:** Leskovec, Kleinberg and Faloutsos (2007) discovered that complex networks densify over time, with the number of edges growing super-linearly with the number of nodes, meaning that the average degree (i.e., number of edges) of the nodes is increasing. In fact, the densification follows a power-law pattern: $e(t) = n(t)^{\alpha}$, where $e(t)$ and $n(t)$ are the number of edges and nodes at time $t$, respectively, and $\alpha$ is an exponent that lies between 1 and 2 ($\alpha = 1$ corresponds to a constant average degree over time, while $\alpha = 2$ corresponds to a very dense graph where, on average, each node has edges to a constant fraction of all nodes). We use exponent $\alpha$ to differentiate the “speed” by which networks are densified.

**Clustering coefficient** measures the probability that two nodes are connected if they already have a common neighbor:

$$C = \frac{3 \times \text{number of triangles in the graph}}{\text{number of connected triples of vertices in the graph}}$$

Intuitively, during the first phase of development, the clustering coefficient of the network is low, since nodes are unconnected with each other. In the second phase, the clustering coefficient tends to increase very quickly as nodes are clustered into very dense, yet unconnected components. When the unconnected components subsequently start to connect with each other, the clustering coefficient drops and stays relatively stable after some time.

**Betweenness** measures the extent to which a particular node lies between the other nodes in the network:

$$B(u) = \sum_{i,j \in E} \frac{\sigma^u(i,j)}{\sigma(i,j)}$$

where $B(u)$ is the betweenness of node $u$, $\sigma^u(i,j)$ is the number of shortest paths between nodes $i$ and $j$ that pass through $u$, and $\sigma(i,j)$ is the number of shortest paths between nodes $i$ and $j$. Nodes with high betweenness have more power to control the information flow in the network, and are normally the gatekeepers who connect several dense groups. For the overall network, the maximum betweenness of all authors is therefore a good indicator of whether there are strong gatekeepers within the network. During the first two phases of the development process, the maximum betweenness is very low, since the nodes are either completely unconnected or clustered in very dense, yet unconnected groups (i.e., there are no controllers in the network). Maximum betweenness increases when more components become connected (emergence stage) and continues to increase when the network develops toward a hierarchical or interdisciplinary topology. However, maximum betweenness will achieve a stable value when the network is at focused stage.

**Largest connected component** (or giant component) measures the fraction of nodes that are connected with each other in the largest sub-network. As observed in Figure 1, this fraction is small in the first two phases, and gradually increases as the network develops and authors from different sub-networks connect with each other. It achieves a stable state when the fraction of nodes that connect to the largest component is equal to the fraction of new nodes that stay unconnected from the largest component.

**Diameter** is the length of the greatest geodesic distance (i.e., the length of the longest shortest path) between any two nodes. Intuitively, in the beginning, the diameter is small, and then it increases. After some time, the diameter starts to shrink as new edges between existing nodes continue to be added. Note that the shrinking of the diameter is not
caused by the emergence of the giant component (Leskovec, Kleinberg & Faloutsos, 2007). However, in our model, if the network develops toward a tree-like topology (hierarchical stage), the diameter will be larger than in the focused and interdisciplinary topologies, respectively.

Average path length is the average length of all the shortest paths in the network. Clearly, during the first two phases, the average path length is small and increases when the network grows. Although communities of conferences and journals are not random networks — and the average path length should therefore be rather small (around six) — there is a slight difference between focused, interdisciplinary and hierarchical topologies. In general, the average path length of a hierarchical network is larger than that of the other two topologies, which gives us more evidence to differentiate these topologies.

In summary, for the co-authorship network, the emergence of the giant component (largest connected component) indicates the cohesiveness of collaboration within the community, while the betweenness shows the existence of the gatekeepers and their importance. The clustering coefficient measures the extent to which the community is clustered into sub-communities. Other parameters such as diameter and average shortest path length, show whether the community is still developing or whether it is stable. For the citation network, combining these parameters helps to understand the interdisciplinarity of a conference.

Data set

The data set used in our study is the combination of DBLP and CiteSeerX digital libraries. DBLP is a computer science bibliography, which also includes publications in interdisciplinary areas of computer science, including TEL. Additionally the whole database is available for download from the website, easing access to the data. We retrieved the publication lists of conferences from DBLP. However, DBLP does not record citations. Therefore, we used CiteSeerX to fill the citation list of publications in DBLP. DBLP data, as downloaded in March 2011, consists of 881,730 author’s names, 1,486,411 publications, 2,868 conference series and 839 journals. The CiteSeerX data set includes 9,121,166 publications, 22,735,140 references and over 6 million author names. We combined DBLP and CiteSeerX using the canopy clustering technique (McCallum, Nigam & Ungar, 2000). Overall, the matching algorithm gave us 864,097 pairs of matched publications. From these data sets we extracted the co-authorship and the citation networks for five main conferences in TEL which are frequently visited by authors of the ET&S journal (see Table 1). The co-authorship networks are based on the DBLP data set. The citation networks are based on the papers and citations in the CiteSeerX data set. Basic statistics regarding the size of the co-authorship and citation networks of these five conference series, including their associated workshops proceedings, are given in Table 2.

<table>
<thead>
<tr>
<th>Table 1. Conferences relevant to the ET&amp;S journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conference Series</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>IEEE International Conference on Advanced Learning Technologies</td>
</tr>
<tr>
<td>Artificial Intelligence in Education</td>
</tr>
<tr>
<td>European Conference on Technology Enhanced Learning</td>
</tr>
<tr>
<td>International Conference on Intelligent Tutoring Systems</td>
</tr>
<tr>
<td>International Conference on Web-Based Learning</td>
</tr>
</tbody>
</table>

* … Average number of papers per event in the conference series which were written by authors who have also published in ET&S. Focusing on recent papers, we only considered conference events since 2005 for calculating this metric.

<table>
<thead>
<tr>
<th>Table 2. Co-authorship and citation network statistics for the selected conferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conference Series</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ICALT</td>
</tr>
<tr>
<td>AIED</td>
</tr>
<tr>
<td>ECTEL</td>
</tr>
<tr>
<td>ITS</td>
</tr>
<tr>
<td>ICWL</td>
</tr>
</tbody>
</table>

327
Development pattern of TEL conference communities

Co-authorship networks

All five conference co-authorship networks are complex networks. In the last seven years, these five conferences combined have published papers written by a relatively stable number of around 1,350 authors each year. For illustration, Figure 3 displays the current co-author network for each of the five conferences in thumbnail form.

---

Figure 3. Co-authorship networks of five relevant conferences in 2010

Figure 4 compares the five conferences with respect to the development of the network parameters introduced in the previous section. Figures 4b-f each plot one of the network parameters on the vertical axis, versus the “age” of the conference series on the horizontal axis. The age in this case refers to an ordered series of conference events without taking the actual time interval between events or the point in time of any specific event into account. Age 1 therefore refers the first event in each series, e.g., ECTEL 2006 and ICWL 2002. Age 2 refers to the second event (e.g., ECTEL 2007 and ICWL 2003), and so forth. Since some conference series have a longer history than others, the data series plotted in Figure 4 have a different number of data points. Of course, to compensate for this imbalance, the number of plotted events could be adjusted to the youngest conference series. However, this would impede the comparative analysis of the development patterns of communities of different conference series: to allow for a fair comparison between two conference series from the perspective of a full life-cycle model (Figure 1), we need to compare the community development starting with the inaugural event for each conference series, at age = 1.

The density — i.e., the ratio between the number of edges and the number of nodes — has increased over time with a coefficient larger than 1 and less than 2 for all five conference series (Figure 4a), whereby the coefficient is largest for ITS (1.38) and ECTEL (1.24), and smallest for ICWL (1.05). This means that ITS and ECTEL successfully manage to match the growing set of authors with a growing web of co-authorship connections. ICWL does not perform well in this regard, which is also evident in the plethora of un- or weakly connected small author groups and the absence of a giant component in the network (see Figure 3) even after nine conference events with 425 published papers.

The clustering coefficient (Figure 4b) of all co-authorship networks is quite high (roughly between .87 and .92) although it has dropped over the years, but Figure 4d shows that AIED and ECTEL have quickly growing largest connected components (i.e., the core author group described for the development model above) indicating a faster scientific community building process than for ICALT and ICWL. ITS has the largest core author group of all five conferences, but it needed longer to develop, since the size remained at an almost constant low value for the first three ITS events. This might be due to the fact that the first three ITS events indexed on DBLP were held between 1992 and 1998, a time span in which annual conferences would have had seven events. While the plots in Figure 4 do not align the actual points in time of different conference series, it seems safe to assume that the time interval between two consecutive conference events, and the overall temporal continuity of conference events, do have an impact on the development of the conference community. More research is needed to clarify this issue.

For maximum betweenness, ITS also has the highest value at slightly under .08 (Figure 4c), which means that the most central author in the ITS network is on almost 8% of all shortest paths through the co-authorship network. This value indicates that there are many active key members — i.e., those authors that connect different author communities through co-authoring of papers — contributing to the conference and community development. ICALT and ICWL do not exhibit such a clear pattern, while AIED and ECTEL are developing very fast in this regard. Fast development of the community typically indicates that the conference has a tighter focus and/or the authors
publishing at those conferences already had strong ties between each other before the conference series started. ECTEL, for example, is a European conference, so the community is by definition smaller than that of ICALT or ICWL, which address TEL communities worldwide.

All diameters of the co-authorship networks are still growing (Figure 4e), indicating that the development of the community is not yet finished. Since the diameter represents the length of the longest shortest path through the network, a peak in diameter growth would indicate a lack of assimilating new author groups into the core conference community. Also, the average path length is still growing for all five conference series (Figure 4f), indicating again that their networks are still growing.

To summarize, ICALT and ECTEL have well-connected authors. The remarkable achievement of ECTEL, which is only half as old as ICALT, is likely based in its origins: the conference was started as an initiative out of the EU project PROLEARN in 2006, and to this day remains a strongly EU TEL project-focused presentation outlet and meeting venue. ICWL, on the other hand, is as old as ICALT, and still seems to struggle with managing the transition from the emergence stage to more mature stages in the development pattern shown in Figure 1. The other two conferences — AIED and ITS — exhibit very mature author communities, which is probably due to the fact that these two conferences attract a strong core of artificial intelligence (AI) researchers. In that sense, they are actually difficult to compare with the other TEL conferences, since their core topic is AI rather than TEL.

Citation networks

The citation networks of the five TEL conferences are complex ones with the ratio between the number of edges and the number of nodes still growing (greater than 1 and less than 2 in Figure 5a). The clustering coefficients of all conferences are similar, with ICWL exhibiting a higher coefficient than the other four conferences (Figure 5b).
However, Figure 5d shows that the literature of ICWL and ICALT is much less connected than that of ITS, AIED and ECTEL, which indicates that the two former ones have a broader, more interdisciplinary scope than the three latter ones. This is supported by the development of the maximum betweenness values in Figure 5c, which indicate the existence of more common core references in these scientific communities. The diameters of ECTEL and AIED have begun to shrink very early, indicating that the body of literature of these communities is quite stable and the themes of the communities are settled. The development of the average path length also supports this finding.

Benchmarking TEL conference communities

We compare the development pattern of TEL conferences to the development pattern of four established conferences in database research, including VLDB (Very Large Data Bases Conference), SIGMOD (ACM Conference on Management of Data), PODS (Symposium on Principles of Database Systems) and ICDE (International Conference on Data Engineering). Because of the long history, outstanding reputation and success of the database research community, these conferences can serve as benchmarks or good practice for other conference communities. The network parameters of these conferences over time are given in Figure 6. All four conferences exhibit the same development pattern: they developed steadily from the bonding stage to a focused topology over a timespan of roughly 20 years. After that time, they achieved a stable network, as we see in the fairly stable values of the clustering coefficient, maximum betweenness, diameter, and average shortest path length.

Compared to the development pattern of TEL conferences in Figure 4, we can see that TEL conferences exhibit a pattern typical of “young” communities. Some TEL conferences develop faster, e.g., ITS, AIED and ECTEL, where betweenness and the largest connected component increase very fast, while the clustering coefficient drops and tends to become stable very early. The other two conferences, ICALT and ICWL, have developed more slowly, but they still follow the same pattern. A closer look at the values of the network parameters shows that ICWL (and, to a
certain extent, ICALT) clearly faces a challenge: the ICWL community is highly clustered into many unconnected components, thus a giant component (a group of core authors) is missing. In this community, we can see the absence of gatekeepers who connect different groups, as indicated by the low maximum betweenness value.

Given the comparison of network structures of TEL and database conference communities, and in particular the case of ICWL, we attempt to understand the strategy by which conferences develop their community. In particular, we are interested in the reason behind the emergence — or the absence — of the giant component. One can imagine two reasons for the absence of such a component:

1. Authors are leaving the conference: if authors publish in the conference once and never come back, they will leave behind “dead” nodes in the co-authorship network, in that they are not active anymore. There will be no connections from these nodes to other nodes in the future. Therefore, new nodes have no chance of connecting to existing nodes.

2. Authors do return to the conference, but they continue collaborating within their own group. This behavior strengthens the connections within groups, but makes no new connections that cross the sub-communities. The whole community is therefore a set of unconnected groups, which contradicts the very nature of a scientific conference.

The giant component is formed when authors choose to stay with the community and collaborate extensively with other authors. The giant component also becomes bigger when new authors are connected to authors who are already
in that component. In both cases, recurring authors play an important role in the development of the community: they ensure the connectivity of the community and their interaction makes the community more cohesive.

To measure the return of authors and their contribution to the community, we calculated the rate of recurring authors and their publications over the years. A paper is published by recurring authors if at least one of the paper’s authors has published in the conference before. A high rate of recurring authors, together with a low rate of papers by recurring authors, indicates that recurring authors mainly collaborate with each other (one paper has more recurring authors). On the other hand, a high rate of recurring authors, together with a high rate of papers by recurring authors, indicates that recurring authors collaborate mainly with new authors, which contributes to community development.

![Graphs showing the frequency of recurring authors and papers over time for TEL and database conferences.](image)

*Figure 7. Recurring authors and papers in TEL and database conferences*

We compared these two measures for the TEL and database research communities (see the plots in Figure 7). As observed, the basic trend during the early stage of the development process is to retain authors: the fraction of recurring authors in all conferences increased in the first years. The frequency of papers by recurring authors also increased. However, keeping this strategy would lead to a closed community and there would be no new ideas coming in from new authors. Therefore, at a certain point in time, conferences retain a healthy fraction of recurring authors. That is one principal strategy for cultivating the scientific community of practice proposed by Kienle and Wessner (2005; 2006). For database communities, we observed this trend in the first 15 years after which the two measures became stable. For TEL conferences, the fraction of recurring authors and their publications in ICWL ceased to increase during the last five years. Year by year, only a small fraction of authors continued to publish in ICWL (less than 25%), while in other TEL conferences, more than 35% authors continued to publish, and this value is still growing. In both TEL and database areas, some conferences quickly managed to retain their authors, e.g., ECTEL and AIED for the TEL community, and PODS for the database community. ICALT currently has the highest...
fraction of recurring authors, but the fraction of papers by recurring authors is less than at ITS, and this rate has developed comparatively slower during the first couple of events.

Depending on the nature of the conference, there are several reasonable explanations for the above observations. Extremely focused conferences such as AIED can quickly manage to retain a good fraction of their authors since there are not so many prominent options when it comes to publishing in this highly focused field. In more interdisciplinary conferences like ICWL, authors return to the conference at a lower rate. Other practical factors will also have an impact, e.g., the location of the conference venue, the programme committee members, and similar factors. For example, ECTEL until now has been held exclusively in Europe, while PODS was held exclusively in North America in the first 22 years. On the other hand, ICALT has moved across the globe from the beginning, and still manages to retain its authors at a high rate. This suggests that a combination of multiple factors determines how well conference communities manage to keep their members returning.

**Discussion and conclusion**

In this paper we have explored the structural development of the Technology Enhanced Learning (TEL) scientific community by analyzing the development pattern of co-authorship networks of five major international conference series in TEL which have particular relevance for authors who have also published papers in the ET&S journal: ICALT, ICWL, ECTEL, ITS and AIED. Co-authorship of a research paper is the most explicit demonstration of ongoing and completed collaborative research in a scientific community of practice, and therefore a valuable object of study in this regard. While we used the development of co-authorship networks (in combination with the paper citation networks) as a main factor for explaining the development of conference communities, we acknowledge that there are many additional factors for conference community development such as the reputation of the speakers and audience attending the conference, the quality of papers at previous conferences, the attractiveness of the keynote speakers and co-located events, the location of the conference venue, the organization skills and community-connectedness of the conference chairs, and many more.

We have calculated and compared social network parameters of the conferences series’ co-authorship networks by applying a time series analysis to reveal patterns and differences in the community development of these five conference series over time. Overall, all five conference series have developed constantly, though at a different pace. Comparing this pattern with that of established conferences in other sub-disciplines of computer science such as databases, we found that TEL conferences exhibit a development pattern that is typical of young and emerging conference communities. Nevertheless, we see that conferences in TEL are building their community in a way that shapes a clear core. In this sense, maintaining and promoting key members who play the role of gatekeepers and connectors is very important.

We have presented a development model for conference communities, including a sequence of phases that may eventually lead to different co-authorship network topologies. We have shown that highly specialized and focused conferences in computer science tend to develop a focused topology with a very large connected set of authors — i.e., a giant component which, for some conferences, may consist of two-thirds or more of all authors in the conference series’ history (e.g., VLDB or ACM SIGMOD). One key feature of TEL, as opposed to highly focused conferences, is its interdisciplinary nature. In the analysis of conference community development patterns, it became evident that interdisciplinarity comes with pros and cons. On the one hand, it attracts researchers from different subject areas to a conference. On the other hand, it slows down the process of building a core group of authors, as we saw, for example, in the relatively slow development pattern of ICWL compared to faster developing, more focused conferences like AIED and ITS, which have a clear artificial intelligence focus, or ECTEL, which has its geographic and thematic focus in European TEL. By far the most important conference series for ET&S authors is ICALT: in the events since 2005, ICALT has attracted an average of over one hundred papers annually by authors who have also published in the ET&S journal.

An analysis of continuity of authorship, i.e., authors who publish in more than one event in the conference series, shows that in the early stages, conferences build their community by retaining authors from previous events. The returning authors and their contributions (papers) are a key driver for the development of a large, well-connected core author group that is characteristic of mature conference communities. However, pushing too hard in this direction would close the community, and there would be no more new ideas coming in in the form of work by new
authors. Conference organizers thus need to strike a balance between measures to retain authors on the one hand, and measures to attract new authors on the other. One potential way of achieving this is to move the conference to a different place (e.g., another continent) every once in a while, since this will attract new, local researchers plus a share of the conference’s veterans. The key success factor appears to be the integration of the established conference community with the newly attracted authors. If we look at ICWL, for instance, there is no clear sign of a large co-author group in the network although the conference has recently celebrated its tenth anniversary. One explanation could be that the conference locations started alternating between Europe and Asia after six venues exclusively located in the Asia-Pacific region, thus posing the challenge of connecting two geographically separated communities after the conference moved to Europe for the first time. This may take more time than expected. On the other hand, ICALT has moved across the globe from the very beginning, and still it managed, during the same time span, to attract more tightly connected authors. This indicates that new authors have been successfully assimilated by the ICALT community.

Understanding the community helps the members to define strategies to support its development. One main goal is to move the community toward a focused topology of connections which will offer a fertile scientific environment for research collaboration. Drawing from the analyses in this paper, some recommendations to different conference community stakeholders for contributing toward this goal are summarized in the following paragraphs.

For conference organizers, besides managing organizational issues such as maintaining stable and reputable committees, or moving the conference to opportune locations to attract and involve local researchers, their efforts in retaining the key authors of the conference are very important for the development of the community. Key authors can be rewarded and attracted in several ways, e.g., through offering them roles in the organizing committees or through opportunities for plenary addresses (e.g., keynotes) or similar occasions where they can spark future cooperation by sharing their work and vision.

The key members of the community, i.e., those with a high centrality in the co-authorship network, can contribute to community development not only by publishing papers, but also by cultivating communication between current and prospective community members. With their knowledge of the conference topics and community, they should be active in finding, suggesting and setting up new collaborations with members in different sub-communities, particularly from the conference network’s periphery, which will make the whole community more integrated and cohesive. Normally, key members are positioned at the interface between sub-communities, so they are aware of the information and ideas emerging from different sources. Subsequently they can synthesize these sources of information to generate ideas and gather together authors from different sub-groups to work on these ideas. Key members also play an important role in engaging new authors and connecting them to the core of the community. This can also lead to the introduction of new ideas and research topics to the conference.

Finally, for all other conference community members, continuing work in their established co-author sub-community will strengthen existing collaboration ties, but may impede the development of the whole conference community. Engaging in collaborations that span different sub-communities helps to generate collaboration ties within the conference community, and can also strengthen the reputation of authors who are sparking these collaborations. In terms of community topology, this also contributes to making the community more cohesive and focused. Additionally, connecting different sub-communities is an important indicator for the future reputation of these authors and their status as “gatekeepers” based on their centrality in the network. In future work we are planning to augment the findings drawn from structural analysis of the conference communities with semantic analysis of papers published at conferences by different authors and author groups. With this structural-semantic analysis we expect to be able to recommend authors for collaboration and papers for reading to community members. It will also provide more insight into the development of the thematic focus of conferences and the roles of key authors.

Acknowledgment

This work has been supported by the BIT Research School of RWTH Aachen University and University of Bonn, and by the EU FP7 projects TEL-Map (http://telmap.org) and ROLE (http://role-project.eu).
References


Flexible Pedagogy, Flexible Practice: Notes from the Trenches of Distance Education
(Book Review)

Reviewer:
Dermod Madden
Associate-Superintendent, Aspen View School Division
EDDE student, Athabasca University
derm.madden@aspeview.org

Textbook Details:
Flexible Pedagogy, Flexible Practice: Notes from the Trenches of Distance Education
Edited by Elizabeth Burge, Chere Campbell Gibson and Terry Gibson
2011, 348 pages, AU Press
ISBN 978-926836201(print) 978-926836218(PDF) 978-926836621(epub)

‘Flexible Pedagogy, Flexible Practice’ is a compilation of essays written specifically for this book, by learned postsecondary educators, from around the world, on the topic of flexibility in post-secondary education; with a particular focus on distance education. The central theme of the book and of each of the twenty-three articles is ‘flexibility’ in learning and in education. Each of the twenty-three articles provides a unique perspective of flexibility in distance education, as defined within the context of a particular country, political system, nationality or cultural group. The final product is a collection of distinctive, yet interconnected perspectives on the relative importance of flexibility in education.

The rationale for the compilation is to create a shared perspective on flexible learning and in so doing, to highlight the challenges faced by educators who advocate for flexibility in education. A second rationale given for the book is, to determine the utility for and relevance of flexible learning as a canonical concept, from a global perspective. This is partly achieved by presenting a broad range of candid reflections from a diverse group of experienced international educators.

Thirty-five authors representing nine different countries and a broad range of educational perspectives contributed to this compilation. The book is divided into five distinct sections. Section one consists of three essays which explore the concept of flexibility. Section two presents four essays on the challenges, referred to in the book as the opposing forces of flexibility, which exist within postsecondary institutions. Section three consists of actual accounts of seven projects which focus on the challenges, successes and failures of implementing change and flexibility in postsecondary education. Section four contains three essays which highlight the pragmatism of change, the realities of compromise and the trade-offs inherent in affecting systemic change. Section five introduces a balance to the anthology by presenting three perspectives which both challenge and caution those who advocate for flexibility in education. The final essay is an introspective reflection on the significance of flexibility for distance education and lifelong learning. The book concludes with a summary of the issues discussed throughout the book and the future possibilities for dialogue on the topic of flexibility.

The strength of this book lies not within the individual efforts of the thirty-five contributing authors and the three editors. Rather, it is the diversity of opinion and perspective presented that illustrates the importance of flexibility as a relative term, dependent on many factors. As such, a definition of and utility for flexibility as a concept, is situational and context dependent.

This is an important book for distance educators with an interest in international education and the challenges of distance delivery from a global perspective. As you read through each of the articles a single theme begins to emerge unmistakable. Flexibility means different things to different people. The concept of flexibility is evolving and changing at a fairly rapid pace, fueled by recent innovations in technology and the resultant changes to government service delivery, industry, and education and society as a whole. Flexibility is context dependent and, as the articles in the book demonstrate, it is also strongly affected by geography, language, culture, political ideology, economics and the need for international partnerships and consensus at a global level. As we move ever closer to a global
perspective, the need for societal flexibility and adaptability becomes more significant. The relevance of this book lies in the fact that it attempts to illustrate the complex contextual and variable nature of flexibility. This is in many ways a book about the subjectivity and variability of flexibility in education.

This book also has relevance for education students at the undergraduate and graduate level. It has significance for studies in educational leadership, instructional design, learning theory, and inclusive education practice. The layout of the book is such that it simulates an interactive discussion about the challenges of flexible delivery in education. I would recommend this book for anyone involved in educational leadership in K to 12 and postsecondary education.

It is not within the mandate of this review to provide chapter summaries for each of the 23 chapters or to delineate the numerous definitions and interpretations of the term ‘flexibility’ found within it. Each chapter narrates different story about the need to create flexible educational environments within a given context. This review will focus on one article to illustrate the importance of context, and the relative nature of specific circumstances.

The title chosen is, ‘Cultural Perceptions of Flexibility in Asian Higher Education’ by Colin Latchem and Insung Jung. The article opens with a retelling of the Indian folktale, ‘The Six Blind Men of Indostan, as retold by American poet, John Godfrey Saxe:

\begin{quote}
It was six men of Indostan, to learning much inclined, who went to see the Elephant (Though all of them were blind), that each by observation might satisfy his mind
And so these men of Indostan, disputed loud and long, each in his own opinion
Exceeding stiff and strong, though each was partly in the right, and all were in the wrong
So oft in theologic wars, the disputants, I ween, rail on in utter ignorance
Of what each other mean, and prate about an Elephant, not one of them has seen
\end{quote}

Latchem and Jung define flexibility as a willingness to change and the ability to bend without breaking. The article focuses on the challenges of education and the need for flexibility in Asian universities. Asian postsecondary education operates within an international knowledge network and as such is impacted by developments and changes in western educational philosophy and practice. They are also tasked with the responsibility of reflecting and celebrating cultural nuances, local traditions and societal norms. Flexibility in this context means different things to different people.

The social, political and economic realities in many Asian countries mandate a degree of flexibility that allows rapid expansion for education while constraining costs; providing post-secondary education to two-thirds of the world’s population, with less than 3% of the world’s wealth. This disparity necessitates a context for flexibility.

In many Asian countries flexibility can also be defined in terms of a movement away from government control. Government controls over universities often require negotiations and lobbying for changes in legislation. This will impact the University's ability to be flexible. Culture can also impact flexibility, especially those aspects of social engagement that are culturally sensitive. Asian universities incorporate aspects of their particular traditions, cultures and social realities. In such contexts, flexibility can mean different things for different people. The authors state that flexibility may mean, in the case of India’s Indira Gandhi National Open University and the Open University of Sri Lanka, the need for open admissions for students; while at another university such as the Open University of Israel, it may mean allowing students the freedom to design their own programs of study.

Limited access to technology can also be a deterrent of flexibility. Internet penetration rates in East, South and central Asia in terms of a percentage of the population is 14%, compared to 22% for the rest of the planet. The authors conclude that every culture should define flexibility within its own philosophical, theoretical and operational framework, while at the same time allowing the flexibility required to manage sensitive cross-cultural issues; and avoid the conclusions of the six men of Indostan.

The significance of this compilation of essays is that it is the first step in a process to create a much needed dialogue on the issue of flexibility in education, with the intent of establishing standards for universal access to education.