Lecture Notes in Educational Technology

Maiga Chang
Yanyan Li *Editors*

Smart Learning Environments

Springer

Lecture Notes in Educational Technology

Series editors

Lecture Notes in Educational Technology

The series *Lecture Notes in Educational Technology* (LNET), has established itself as a medium for the publication of new developments in the research and practice of educational policy, pedagogy, learning science, learning environment, learning resources etc. in information and knowledge age, – quickly, informally, and at a high level.

More information about this series at http://www.springer.com/series/11777

Maiga Chang · Yanyan Li
Editors

Smart Learning Environments

Springer

Editors
Maiga Chang
Athabasca University
Edmonton, AB
Canada

Yanyan Li
School of Educational Technology
Beijing Normal University
Beijing
China

ISSN 2196-4963 ISSN 2196-4971 (electronic)
ISBN 978-3-662-44446-7 ISBN 978-3-662-44447-4 (eBook)
DOI 10.1007/978-3-662-44447-4

Library of Congress Control Number: 2014946402

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The term Smart Learning Environments is relatively new in educational technology. And while there is no agreed upon meaning as yet, the diverse papers in this fine volume provide a fresh vantage point for viewing the road ahead in research and new product development.

Since Kuhn's *Structure of Scientific Revolutions*, we are used to thinking of science as progressing in leaps, not increments. And leaps occur through unification across discipline boundaries. Smart Learning Environments, and the results reported in this study, cluster around three broad themes: *real-time analytics*, *ambient design*, and *smart pedagogy*. The next wave of innovation in research and product development is likely to emerge at points of intersection and synthesis from these three areas.

The first theme is learning analytics. Early research contributed to developing traditional dashboards, but this has largely provided retrospective data about learners. Since then predictive models have targeted identifying at-risk students preemptively. A number of papers in this volume suggest that the next leap will come from *real-time* analytic systems that provide *just-in-time* feedback in the *learning moment* to learners and instructors. For example, Chap. 1 describes how we might infer Working Memory Capacity (WMC) from the real-time logs of learning systems and then pass the inferences as streams to adaptive systems. Similarly, Chap. 2 explores how we might establish causal relationships between student activities and learning outcomes by real-time tracking of coding practices of computer science students. "Real-time" analytics, by linking the *end state* of learning with a deep understanding of *how* students got there, should enable advanced mechanisms for feedback in next-generation learning systems.

The second theme is ambient design. Products are too often seduced by technology and technologists. Ambient design fits technology to the learning experience, not the reverse. There is considerable evidence, reaching back to Seymour Papert's pioneering work, which suggests that fieldwork is one of the best means to stimulate exploratory learning. Drawing on this hypothesis, Chap. 5 designs an outdoor *exploratory learning experience* by harnessing mobile and location-aware technology. Their EagleEye platform incorporates student-centered learning while also equipping the instructor with rich scaffolding tools for feedback, exploration, and collaboration. Chapter 6 provides ethnographic evidence that in China student note-taking in traditional textbooks serves as a focal point for family-school communication. Although this practice is apparently unique in China, their research suggests important *design principles* for creating next-generation smart books.

The third theme is smart pedagogy. In a provocative paper, Chap. 8 warns of the dangers of automation isolated from the school routine, such as in Intelligent Tutoring Systems. They go on to sketch a framework where teachers work in *tandem* with technology to provide constant quality feed to students. Their model incorporates the *judgment* of teachers alongside learning data from analytic systems to provide the most accurate *smart feedback* to students. Chapter 9 describes a motion-based exercise game intended to train balance and improve physical health in older adults. As has been found in previous studies, the participants enjoyed playing with others rather than alone. Even in late stages of life it is seen that *communal discovery* and social interaction improve motivation and learning outcomes.

The papers in this rich volume forge a new understanding of the interplay between pedagogy and technology. Researchers and students will be amply rewarded by a careful study of *Smart Learning Environments*.

Alfred Essa
Vice President, R&D and Analytics, Digital Platform Group
McGraw-Hill Education

Preface

This edited volume consists of selected papers from distinguished experts and professors in the field of smart learning environments. This book addresses the main issues concerned with the future learning, learning and academic analytics, virtual world and smart user interface, and mobile learning. The learning environment has been affected by advances in technology development and changes in the field of education. What we should do is to make the learning systems, no matter what platforms (i.e., personal computers, smartphones, and tablets) they are running at, be aware of the preferences and needs that their users (i.e., the learners and teachers) have and are capable of providing their users the most appropriate services and help to enhance the users' learning experiences and to make the learning efficient. This book aims to gather the newest research results of smart learning environments from the aspects of learning, pedagogies, and technologies in learning.

This book arranges research based on three themes: learning analytics, ambient design, and smart pedagogy. Each chapter covers three to four latest research results related to the development of smart learning environment for the future learning. The aim is to provide readers with evidence and experiments that account for users' experiences and perceptions related to knowledge and concepts acquisition through these smart learning approaches in various disciplines and domains.

First, this book starts with Dr. El-Bishouty and his colleagues' three research results that automatically analyze both learners' characteristics and courses in learning systems based on learners' cognitive abilities, learning styles, and course context. The first research result is a system, which can automatically identify students' working memory capacity based on their behavior, is developed. By knowing students' working memory capacity, teachers can provide students personalized support and recommendations so the students can learn better. The second research result is a tool that can analyze course contents and check whether or not the current course contents are suitable for the students who enroll in this course. Teachers may adjust course contents by adding, moving, or removing learning activities to make the course contents more appreciated by the students.

The third research result is an application that can build a comprehensive context profile through detecting available features of a device and tracking the usage of these features. The context profile can help researchers to design advanced adaptive and intelligent capabilities for their learning systems and environment.

The second chapter talks of an approach that can measure students' coding competency proposed by Dr. Kumar and his colleagues. They first talk of the types of learning traces that they collect from the Java tutor sensor (i.e., MILA) which students install in their Eclipse development environment before they start doing programming assignments. Dr. Kumar and colleagues then explain the data analysis methodology including the creation of key performance indicators for the three dimensions: overall course measurement, student-based traces, and content-specific traces. The technologies they currently include are simple statistics, rule-based approaches, pattern recognition, mixed-initiative conversational agents, and causal models. Their research can discover causal connections between study pathways and competencies.

Dr. Gao and Dr. Wen in the third chapter propose an enhanced semantic similarity topic modeling method for document analysis. They not only use co-occurrence information but also the semantic similarity based on WordNet as auxiliary information while analyzing the similarity among documents. They evaluate the proposed method by comparing it with existing topic modeling methods and find the accuracy that the proposed method reach is around 1.26–1.31 times higher than the existing methods. As the learning repositories grow rapidly due to the easy of publishing and accessing information, the proposed method can help students to understand the knowledge in documents and organize documents quickly.

The last chapter in the learning analytics theme is Dr. Li and her colleagues' effective approach in identifying and recommending experts with high expertise and influence in online learning communities. Their method not only considers the topic similarity degree that the documents posted by users have, but also measures the quality of the documents based on user feedback, review sentiment, and topic-specific influence degree of the users who give feedback. Dr. Li and colleagues use the gold standard method to evaluate the proposed approach's accuracy and find the correlation coefficient value to be sufficiently high. Their research can be applied to large size of learning community such as MOOCs. In the large learning community, many produced artifacts, postings, and learners are loosely tied and the proposed approach can be used to identify and recommend experts with high expertise and influence to make an online learning community interconnected and cohesive.

The ambient design theme starts with Dr. Jong's Chap. 5. In Chap. 5, Dr. Jong explains a GPS-supported integrated educational system, EagleEye, which can help both teachers and students to do context-aware outdoor exploratory Geography learning activities. He finds that EagleEye not only better promotes collaboration among students in terms of doing learning activity in the fieldtrip, but also earns teachers' positive perceptions toward its educational innovation—student-centeredness, motivation, scaffolding, and user-friendliness.

Dr. Chen and his colleagues in Chap. 6 introduce their investigation and findings on the note-taking that pupils have on the textbook. First, they summarize a note's three features (i.e., forms, locations, and the contents) and the note-taking influence factors (i.e., student's initiative, teacher's lecturing speed, and the content's difficulty). They also find that the contents of a note are quite different from the textbook's discipline and the student's age. Their findings are important for researchers in the e-text area. Researchers have to consider the features and functions (i.e., a mediator for family-school communication) while developing e-text readers and content.

Chapter 7, the last chapter in the ambient design theme, is written by Dr. Hu and Dr. Huang. They talk of the effectiveness that Clickers—a small device that allows students to quickly answer questions presented in the class—have if compared with peer discussion, for secondary school students learn biology. They find that both Clickers and peer discussion strategies are good for students to gain knowledge of biology and there is no significant difference in terms of students' learning achievements. Although Clickers may get students motivated in learning than peer discussion strategy, there is no significant difference in terms of students' learning attitudes among the two strategies.

The third theme—the smart pedagogy theme—starts with Dr. Libbrecht and his colleagues' Chap. 8. In their chapter, they argue that a learning tool can be smart if it has semi-automatic feedback paradigm designed. For semi-automatic feedback, their definition is—teacher is able to complement the feedback generated by the learning tool by a relevant feedback. In this chapter, they also introduce two smart learner support through semi-automatic feedback-based learning tools and correspondent use cases. The tasks that teachers are supposed to do are also summarized at the end of this chapter.

Mr. Fang and his colleagues in Chap. 9 introduce a motion-based Evergreen Fitness System that enables older adults' balance training. In order to design such game, they consulted two domain experts to design proper exercises for improving balance performance. They also invite senior participants to evaluate the suitability of the movements and the enjoyment of using the system, as well as provide feedback to improve the prototype. Several important conclusions of developing such system are: (1) such system should include enough in-game navigation, so learning how to use the system is not required; (2) the content should be more relevant to the users' daily life; and (3) such system should be capable of adjusting the motion recognition according to individual differences.

Chapter 10 talks about the use of story-based virtual laboratory for Physics developed by Mr. Fan and his colleagues. The virtual lab can help students to sharpen their science process skills that are usually not given attention in traditional science learning. Their experiment result shows that students, especially female students, believe that the system is useful for them in terms of learning. They also find that students' past physics grades and TIPS (i.e., Test of Integrated Science Process Skills) scores do not influence students' intention of using the system; which means students with low academic achievement can also benefit from the use of the virtual laboratory and improve and practice their science process skills.

The last chapter of the smart pedagogy theme and this book is written by Dr. Santos and his colleagues. They use augmented reality to develop an X-ray application which allows students to see inside of objects with mobile devices like a tablet. Surprisingly, their research results show that augmented reality X-ray does not have a significant impact on students' perceptions toward depth and realism. Via interviews with teachers, they find the augmented reality X-ray is perceived to be useful but a high-quality teaching plan is missing. Moreover, both teachers and students need to be trained to use the augmented reality X-ray tool in the classroom.

Maiga Chang
Yanyan Li

Contents

Part I
Learning Analytics

Chapter 1
Analyzing Learner Characteristics and Courses Based on Cognitive Abilities, Learning Styles, and Context

Moushir M. El-Bishouty, Ting-Wen Chang, Renan Lima, Mohamed B. Thaha, Kinshuk and Sabine Graf

Abstract Student modeling and context modeling play an important role in adaptive and smart learning systems, enabling such systems to provide courses and recommendations that fit students' characteristics and consider their current context. In this chapter, three approaches are presented to automatically analyze learners' characteristics and courses in learning systems based on learners' cognitive abilities, learning styles, and context. First, a framework and a system are presented to automatically identify students' working memory capacity (WMC) based on their behavior in a learning management system. Second, a mechanism and an interactive tool are described for analyzing course contents in learning management systems (LMSs) with respect to students' learning styles. Third, a framework and an application are presented that build a comprehensive context profile through detecting available features of a device and tracking the usage of these features. All three approaches contribute toward building a foundation for providing learners with intelligent, adaptive, and personalized support based on their cognitive abilities, learning styles, and context.

Keywords Cognitive abilities · Learning styles · Context profile · Student modeling · Personalization

---

M.M. El-Bishouty (✉) · T.-W. Chang · R. Lima · M.B. Thaha · Kinshuk · S. Graf
Athabasca University, Athabasca, Canada
e-mail: moushir.elbishouty@athabascau.ca

T.-W. Chang
Beijing Normal University, Beijing, People's Republic of China

M.M. El-Bishouty
City for Scientific Research and Technological Applications, Alexandria, Egypt

R. Lima
Federal University of São Carlos, São Carlos, Brazil

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_1

1.1 Introduction

Learners have different backgrounds, motivation, and preferences in their own learning processes, and learning systems that ignore these differences have difficulty in meeting learners' needs effectively (Atman et al. 2009). Therefore, when designing instructional material, it is important to accommodate elements that reflect individual differences in learning. Learning systems that provide adaptivity and intelligent support based on learners' characteristics and situations have high potential to make online learning and teaching easier, and more effective for both learners and teachers (Tseng et al. 2008; Macfadyen and Dawson 2010). However, most of the popular learning management systems (LMSs) currently used by educational institutions, such as Moodle and Blackboard, do not provide adaptivity or intelligent support, neither for learners nor for teachers.

The aim of our research was to make learning systems more intelligent, adaptive, and personalized by analyzing online course structures and building comprehensive learner and context profiles that include various information about the learner and his/her context. Such analysis results and comprehensive profiles can then be used to extend learning systems with advanced adaptive and intelligent capabilities that provide personalized user interfaces, and adaptive course structures and recommendations. This book chapter focuses on the following three elements: learners' cognitive abilities, learning styles, and context. The approaches described in this chapter aim at gathering and provide information that enable learning systems to be adapt to the learners' cognitive abilities and learning styles, as well as the current context in which the learning occurs, leading to more effective, convenient, and successful learning experience.

One of the most important cognitive abilities for learning is working memory capacity (WMC) which affects students' learning behaviors to perform complex cognitive tasks such as reading comprehension, problem solving, and making decision (Broadway and Engle 2011). Providing students with course materials and activities by considering their WMC helps in avoiding cognitive overload and therefore positively affects students' learning (Gathercole and Alloway 2008). Traditionally, WMC can be measured by a variety of memory span tasks including counting span, operation span, and reading span which are related to the complex cognitive tasks (Broadway and Engle 2011; Unsworth et al. 2012). However, the main disadvantage of these measured tasks is that students have to do them in addition to their learning. In this chapter, we present a framework for automatically identify WMC from students' behavior in a learning system, which is described in Sect. 1.2.

Clay and Orwig (1999) defined learning style as a unique collection of individual skills and preferences that affects how a person perceives, gathers, and processes information. Making teachers aware of how well their courses fit with diverse learning styles can help them in improving their courses and providing personalized support for students with different learning styles. Sect. 1.3 illustrates our mechanism for analyzing course contents in LMSs with respect to learning styles and a visualization tool to show the course support level for students' learning styles.

Data gathered from the usage of the learners' devices provide important information about the user and his/her context and can, together with the availability of device functionalities/features, be used to build a context profile (Chen et al. 2008). In Sect. 1.4, we introduce a framework that aims at building a comprehensive context profile, containing information about the devices a learner uses, the available functionalities/features on a learner's device as well as how frequently the respective device functionalities/features are used by the learner.

Section 1.5 concludes the chapter.

1.2 Identifying Cognitive Abilities from the Log Information of Learning Systems

The cognitive abilities are the abilities of the human brain to understand complex ideas, to adapt effectively to the environment, and to learn from experience (Broadway and Engle 2011; Park and Lee 2003). WMC, one of students' cognitive characteristics, is to keep active a limited amount of information for a very brief period of time (Miller 1956). Results of several studies have shown that students with low or high levels of WMC have very different performances on the different attention levels during performing cognitive tasks (Broadway and Engle 2011; Engle 2010; Gathercole and Alloway 2008). Knowing the levels of students' WMC can help in many ways to enhance learning and teaching in learning systems. Teachers can use this information to provide meaningful recommendations to their students. Furthermore, information about students' WMC can be used as input for adaptive systems to provide students with customized learning content and activities to suit their individual WMC. This section focuses on how to identify students' WMC from their learning behaviors in learning systems. A general framework for automatically identify WMC from students' behavior in a learning system is introduced in this section. The proposed framework is not restricted to a particular learning system and extends our work on identifying learning styles (Graf et al. 2009a) by additionally considering WMC in the detection process.

The framework consists of two components: the data extraction component and the calculation component, as shown in Fig. 1.1. The data extraction component is responsible for extracting relevant data from the learning system's database, preprocessing these data, and passing the preprocessed data to the calculation component. The calculation component then uses these data to detect patterns from students' behavior (in the Pattern Detector) and calculates students' WMC from the detected patterns (in the WMC Calculator).

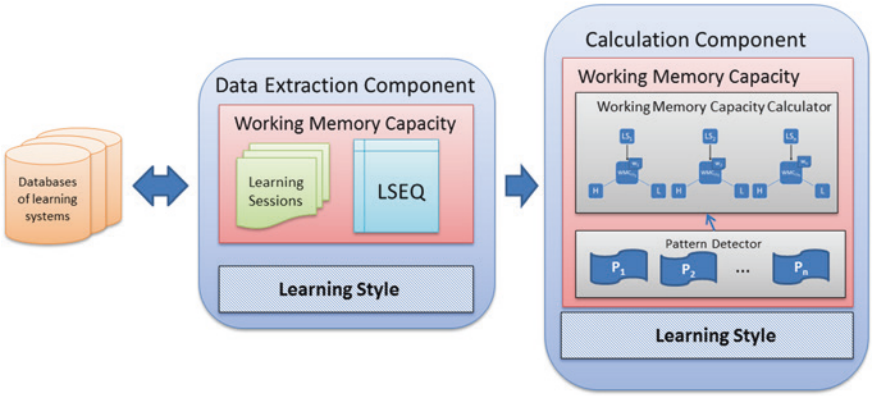

Fig. 1.1 Framework for Identifying WMC

1.2.1 Preprocessing of Data in Data Extraction Component

The detecting tool, DeLeS (Graf et al. 2009a), has been extended in order to identify not only learning styles, but also WMC. In order to analyze students' behavior from the log information in a database of a learning system, some preprocessing of behavior data and course data need to be conducted. The preprocessing includes the following: (1.1) identifying learning sessions, (1.2) filtering out activities that are not dedicated to learning as well as activities where students visit a learning activity only for a very short time, and (1.3) building a learning sequence table called LSEQ table that includes the structure of the course in terms of the predefined sequence of learning activities/objects in a course. The preprocessing of data provides the required input data for the calculation component in order to analyze relevant behaviors for WMC detection. In the next subsections, the three preprocessing steps are described in more detail.

1.2.1.1 Identifying Learning Sessions

A learning session is a series of learning activities that a student does while focusing on learning. A learning session typically starts when a student logs in and ends when a student logs out. Furthermore, we additionally consider breaks in learning or simply closing the learning system without logging out. In order to do so, for each learning activity, an upper threshold is predefined by teachers based on the type of activity to indicate the maximum time a student would typically spend on a learning activity. This threshold is used to identify learning breaks based on log data which typically include timestamps of each activity that a student does. Such learning breaks then indicate the end of one learning session and the beginning of a new learning session.

1.2.1.2 Filtering Learning Activities

When analyzing the behavior patterns of students to infer indications about their WMC, we only focus on learning activities. Therefore, general activities such as a student making modifications to his/her user profile or a user checking his/her marks are not seen as learning activities and are not considered in a learning session. Furthermore, we filter out activities where a learner spent only a very short time (e.g., because he/she clicked on the wrong link or is searching through pages until he/she finds the page that he/she actually wants to read).

1.2.1.3 Building a Learning Sequence Table

We extract the information about the sequence in which the course is laid out and store this information in an internal database table called LSEQ table. When a course designer/teacher makes any changes to the course structure in the learning system, the modification time is stored in the internal database. Once the course designer/teacher uses the DeLeS tool to detect students' WMC, this modification time is used to check whether the LSEQ table needs to be updated before using it to calculate students' WMC.

1.2.2 Analyzing Relevant Behavior Patterns in the Pattern Detector

The pattern detector of the calculation component is responsible for analyzing relevant behavior patterns. Six patterns are considered, which are explained in the subsequent subsections.

1.2.2.1 Linear Navigation Pattern

Linear navigation means that students learn the materials linearly and follow the learning sequence of the course defined by teachers. Huai (2000) performed an experiment to investigate the relationship between WMC, long-term memory, and a serial/holistic learning style. To draw conclusions about the relationship between WMC and a serial/holistic learning style, linear and non-linear navigational behavior of students was investigated. As a result, Huai also found that students with high WMC tend to focus on linear navigation and students with low WMC tend to use non-linear navigation. Accordingly, if such linear navigation is found, it gives an indication for high WMC. Otherwise, if non-linear navigation is found, it gives an indication for low WMC.

1.2.2.2 Constant Reverse Navigation Pattern

Reverse navigation means that a student revisits an already visited learning object (LO). Constant reverse navigation indicates that a student frequently goes back to an already visited LO. This behavior can be explained by the limited capacity of working memory for students with low WMC (Graf et al. 2009b). The process of constant reverse navigation is caused by an insufficient WMC to hold on the materials that have just been visited (Lin et al. 2003). When the learning materials that a student just read on the previous page should be still fresh in his/her working memory, the constant need to navigate backward is a sign of working memory deficiency. The definition of constant reverse navigational behavior is that there are more than two LOs revisited in the same learning session and the navigational relations of these LOs are not defined in the LSEQ table (and therefore not in line with the sequence of LOs in the course structure). For example, consider a navigation sequence of a student's behavior as follows: from LO A, to B, to C, to D, to A, and to C. In this navigation sequence, two LOs A and C are revisited, and consequently constant reverse navigational behavior is detected. If the constant reverse navigational behavior is found, it gives an indication for low WMC. Otherwise, an indication for high WMC is found.

1.2.2.3 Simultaneous Tasks Pattern

The simultaneous tasks pattern is transferred from the ability of attentional control on performing two tasks simultaneously. Previous studies have shown that when performing two tasks simultaneously, low WMC participants were less accurate than participants with high WMC (Engle 2010; Woehrle and Magliano 2012). For identifying this pattern, overlapping navigational behavior is investigated to indicate that a student is trying to perform two tasks, the visiting and evaluation tasks, simultaneously. If a student visits at least one other LO in between LO A (visiting task) and its evaluation *EA* (evaluation task), overlapping navigational behavior is found. In such situation, the student learns LO A first and then learns other LOs before taking the evaluation of LO A. Therefore, he/she needs to remember the concept of LO A in his/her working memory while learning other LOs. If the student then passes the evaluation of LO A, the simultaneous tasks pattern is found, which gives an indication for high WMC. If he/she fails, the non-simultaneous tasks pattern is identified, which gives an indication for low WMC.

1.2.2.4 Recalling Learned Material Pattern

The recalling learned material pattern is transferred from the relationship between WMC and long-term memory. This pattern is similar to the simultaneous tasks pattern, but it is identified within two different learning sessions.

Prior works have argued that the individual's ability to retrieve information from long-term memory is determined by his/her WMC (Unsworth et al. 2012; Engle 2010). As a result, they found that low WMC participants cannot recall as much information from long-term memory as high WMC participants since low WMC individuals do not search the remembered information in their long-term memory as effectively as high WMC individuals. This pattern is found if a student visits LO A in one session but does not perform an evaluation of his/her knowledge on LO A in that session. In a different learning session, the student then does not visit LO A but goes directly to the evaluation of LO *A (EA)*. If the student then passes the evaluation, it means that he/she could recall the previously visited information from LO A from his/her long-term memory and the recalling learned material pattern is found, which gives an indication for high WMC. If he/she fails the evaluation, the non-recalling learned material pattern is identified, which gives an indication for low WMC.

1.2.2.5 Revisiting Passed Learning Object Pattern

As mentioned in the previous sections, several studies have argued that individuals with low WMC cannot recall as much information from long-term memory as individuals with high WMC (Engle 2010; Unsworth et al. 2012). This pattern considers a situation where a student visited LO A and successfully completed its evaluation (EA) in the same session, but then revisits LO A afterward in a different learning session. In such case, the student seems to have problems recalling information from his/her long-term memory and wants to reread some of the already learned information. The more time the student spends on LO A during such revisit, the more problems the student seems to have in recalling the respective information from the long-term memory and therefore, the stronger the indication for low WMC is. On the other hand, if the time the student takes for reading and recalling on LO A is low, an indication for high WMC is given.

1.2.2.6 Learning Style Pattern

The learning style pattern is based on the relationship between learning styles and WMC. Graf et al. (2009b) investigated the direct relationship between WMC and the four learning style dimensions of the Felder-Silverman learning style model (FSLSM) (Felder and Silverman 1988), namely the active/reflective, sensing/intuitive, visual/verbal, and sequential/global dimensions. The results of the study showed that students with a reflective or intuitive learning style tend to have high WMC and students with an active or sensing learning style tend to have low WMC. For the visual/verbal dimension, the study found only a one-directional relationship, namely that learners with a verbal learning style tend to have high WMC, whereas visual learners have either high or low WMC. No relationship for the sequential/global dimension was found. The learning style pattern considers these relationships. Accordingly, if a student has an active or sensing learning style, this gives an indication for low WMC. On the other hand, a reflective, intuitive, or verbal learning style gives an indication for high WMC. An average value of all indications from a student's learning styles is calculated and this value represents the overall indication of WMC for this learning style pattern.

1.2.3 Calculating Student's WMC in WMC Calculator

After preprocessing the data, these data are used to calculate the students' WMC in the WMC calculator based on the five navigational behavior patterns and the learning style pattern. If a navigational behavior pattern is detected in a relation between two LOs, this relation is considered as an activated relation for the particular pattern. In each learning session, a value p is calculated for each of the five navigational behavior patterns based on the number of activated and non-activated relations in this session. This value p shows how strongly the student's behavior represents the respective pattern. Subsequently, the value p for each pattern is transferred to its indication for WMC (e.g., a high p value for linear navigation provides an indication for high WMC). Then, the indications from the five navigational behavior patterns and the indication based on the learning style pattern are summed up and divided by the number of activated patterns (where the learning style pattern is considered as activated as soon as the learning styles of the student are known). The result of this calculation represents the indication for WMC of the respective learning session. Although the learning style pattern is not dependent on learning sessions, we decided to add the indication from this pattern in each session in order to ensure that this pattern has the same impact in the detection process as all other navigational behavior patterns. Each learning session also contains a weight, which determines the influence of each session on the overall value of WMC and is calculated based on the number of activated relations in a session for all patterns. In order to calculate the student's WMC, the WMC indication of each session is multiplied by the weight of the respective session. Subsequently, the results for all sessions are summed up and divided by the number of sessions. The resulting value is the identified WMC for the respective student.

The DeLeS tool for identifying students' WMC and LS based on their behavior has been implemented as a block plugin into the learning management system Moodle. For the visualization of individual student's WMC, as show in Fig. 1.2, a teacher can select a particular student and can see his/her WMC, which is presented on a scale from 0 to 1, where 0 indicates a very low WMC, and 1 indicates a very high WMC.

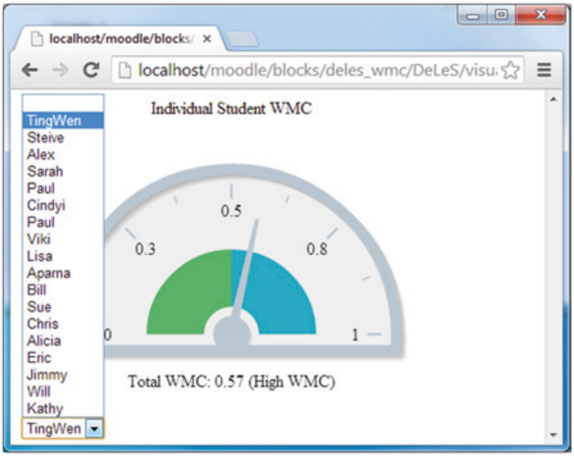

Fig. 1.2 Individual Student WMC

Furthermore, teachers can access more detailed information about students' WMC with respect to a particular learning session, as shown in Fig. 1.3. A teacher can see the students' behavior of each learning session based on the six patterns (LN: Linear Navigation; CR Constant Reverse navigation, ST: Simultaneous Tasks, RC: ReCalling learned materials, RV: ReVisiting passed learning object, and LS: Learning Style) and what indication this behavior gives with respect to a student's WMC. As shown in Fig. 1.3, the blue bars represent indications for high WMC and the green bars represent indications for low WMC. The WMC of each session is calculated based on the indications from all patterns. For example, Fig. 1.3 shows that the respective student performed more linear navigation behavior than non-linear in session 6 and therefore, the LN pattern gives an indication

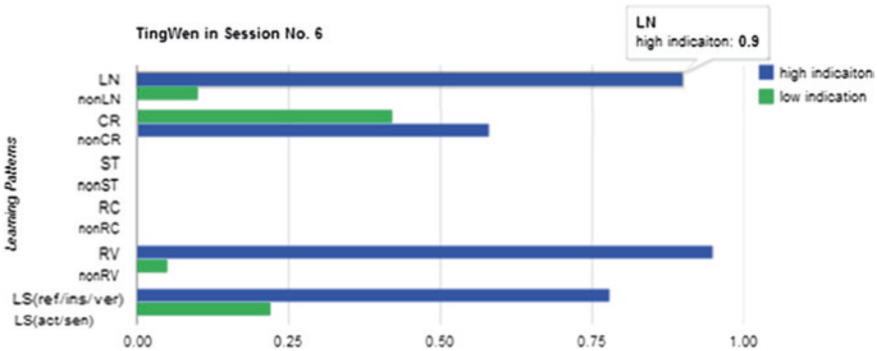

Fig. 1.3 Interface for showing WMC analyzed results in a particular learning session—LN, CR, ST, RC, RV represents the activated behaviors of respective patterns; nonLN, nonCR, nonST, nonRC, nonRV represents the non-activated behaviors of those patterns; LS(ref/ins/ver) represents reflective, intuitive, and verbal learning styles; LS(act/sen) represents active and sensing learning styles

value of 0.9 on a scale from 0 to 1 where 0 represents a strong indication for low WMC and 1 represents a strong indication for high WMC. Accordingly, the indication value of 0.9 suggests that the student has high WMC.

1.3 Analyzing Course Contents in Learning Management Systems with Respect to Learning Styles

A person's learning style is the method that best allows the person to gather and to understand knowledge in a specific manner. Once a learner's particular learning style is identified, it is possible to identify ways in which the learning process can be improved (Onyejegbu and Asor 2011). On the other hand, making teachers aware of how well their courses fit with diverse learning styles can help them in improving their courses to support students with different learning styles.

Our research utilizes the Felder and Silverman's Learning Style Model (FSLSM) (Felder and Silverman 1988) because of its applicability to e-learning and compatibility to the principles of interactive learning systems design (Kuljis and Liu 2005). In this model, Felder and Silverman proposed four dimensions of learning styles (active/reflective, sensing/intuitive, visual/verbal, and sequential/global) and teaching styles (active/passive, concrete/abstract, visual/verbal, and sequential/global), where each teaching style corresponds to (matches with) a learning style. Many researchers have conducted research to detect the learners' learning styles and provide recommendations and adaptations for online courses based on learning styles (e.g., Paredes and Rodríguez 2004; Graf and Kinshuk 2007). Our research is different from the previous works in that we focus on analyzing existing online courses and making teachers aware of how well those courses fit with diverse learning styles and the current cohort of learners. This section illustrates the design and the implementation of a mechanism for analyzing the course contents, and a visualization tool that shows the support level of the course for diverse learning styles.

1.3.1 Course Analyzing Mechanism

A mechanism for analyzing existing courses in LMSs is proposed in order to infer which learning styles they currently support (El-Bishouty et al. 2012). The mechanism recognizes how well each section of an existing course fits to each of the eight poles of FSLSM (Felder and Silverman 1988) by calculating the average of three factors: the availability, the frequency, and the sequence of the learning objects (LOs) in that section. Consequently, the results can be summarized for each section and for the whole course. The mechanism currently considers eleven types of LOs as follows:

- *Commentaries*: provide learners with a brief overview of the section.
- *Content objects*: are used to present the learning material.
- *Reflection quizzes:* include one or more open-ended questions about the content of a section.
- *Self-assessment tests*: include several close-ended questions about the content of a section.
- *Discussion forum activities*: provide learners with the possibility to ask questions and discuss topics with their peers and instructor.
- *Additional reading materials*: provide learners with additional sources for reading about the content of the section.
- *Animations*: demonstrate the concepts of the course in an animated multimedia format.
- *Exercises:* provide learners with an area where they can practice the learned knowledge.
- *Examples*: illustrate the theoretical concepts in a concrete way.
- *Real-life applications*: demonstrate how the learned material can be related to and applied in real-life situations.
- *Conclusions*: summarize the content learned in a section.

Certain LO types can support diverse learning styles; on the other hand, it is possible that they have no effect. The availability of types of LOs is considered as a factor to infer the learning styles that a section of the course fits well. It measures the existence of LO types in the section that can support each learning style (ls) in respect to all types of LOs that can support that learning style. The availability factor is calculated using formula 1.1. On the other hand, the frequency factor measures the number of LOs in the section that support each learning style in respect to the frequency threshold. The frequency threshold represents the sufficient number of LOs in a section to fully support a particular learning style. This threshold is predefined and can be adjusted by the teacher if needed. If the number of LOs that support a particular ls in a section is less than the value of the frequency threshold, then the frequency factor is obtained by formula 1.2, otherwise the frequency factor takes the value 1, which means a full frequency support level for that learning style. The obtained values for both, the availability factor and the frequency factor, range from 0 to 1, where 1 indicates a strong suitability to the learning style and 0 means no support.

$$Ava_{\mathrm{ls}} = \frac{(\#\text{ of existing LO types that support ls})}{(\#\text{ of LO types that support ls})} \tag{1.1}$$

$$Freq_{\mathrm{ls}} = \frac{(\#\text{ of existing LOs that support ls})}{(\text{frequency threshold})} \tag{1.2}$$

$$Seq_{\mathrm{ls}} = \frac{\sum_{i=1}^{n} f_{\mathrm{ls}}(\mathrm{LO}_i) \times w_i}{\sum_{i=1}^{n} w_i}, \quad 0 < w \leq 1 \tag{1.3}$$

Actually, not only the types but also the order and the position of the LOs affect the suitability of a course regarding different learning styles. The sequence factor measures the suitability of the sequence of LOs for different learning styles. We calculate the sequence factor for each LO according to its type, location, and order. It is determined according to how much this LO type in that place fits with each of the eight learning styles of FSLSM. The sequence factor for each learning style is calculated using formula 1.3. In this formula, f_{ls} (LO_i) = 1, if LO_i is suitable for that learning style at that location, and f_{ls} (LO_i) = 0 otherwise. The weight w represents how well the position of a LO fits to the learning style, and n is the number of LOs in the section. Formula 1.3 represented the weighted mean of $f_{ls}(LO_i)$. Its value ranges from 0 to 1, where 1 indicates a strong suitability for the learning style and 0 means no support.

1.3.2 System Implementation

Based on the course analyzing mechanism, a tool is implemented for analyzing and visualizing the suitability of a course for students' learning styles (El-Bishouty et al. 2013). It is called interactive course analyzer. By utilizing drag and drop functionality, the tool allows the teacher to play around with the course structure (by adding, moving, and/or removing LOs) showing the expected changes in the course support level for diverse learning styles. The interactive course analyzer aims at helping teachers to improve the support level of their courses by making efficient modifications in the course structure to meet the need of different students' learning styles.

This tool is implemented as a web application. It is mainly developed using MySQL relational database management system and PHP scripting language. It is a stand-alone application that runs on the server side. It connects to a LMS database (Moodle as an example), retrieves the existing course structure and applies the course analyzing mechanism to analyze the course contents with respect to learning styles.

Moreover, the interactive course analyzer visualizes the analysis results and shows how well a course fits with students' learning styles. It provides the teacher with two visualization modes: the General Mode and the Cohort Mode. The General Mode visualizes the support level of a course for diverse learning styles based on FSLSM. The Cohort Mode visualizes the support level of a course in respective to the learning styles of the cohort of students enrolled in that course. Students' learning styles can be calculated through, for example, the ILS questionnaire (Felder and Soloman 1997) or by a tool such as DeLeS (Graf et al. 2009a).

1.3.3 System Usage

There are two methods to access the interactive course analyzer. The first method is embedded into the LMS interface through an add-on that works for a particular LMS (e.g., a block in Moodle), which provides a direct link to the tool interface. This method automatically retrieves the teacher's authentication information from the LMS, and adapts the interface for the teacher's current course and students enrolled in that course. The second method is a standalone web application that requires entering the authentication information manually (such as the user name and password to access the LMS and the internal databases).

Whenever the teacher accesses the tool to analyze a particular course, the interactive course analyzer retrieves the course structure data, analyzes the course based on learning styles, retrieves the learning style data of students enrolled in that course, and then visualizes the course support level. The visualization part of the tool consists of four similar charts that show how well the course and a selected section fit with students' learning styles. Two of the charts visualize the course support level before and after the modifications made by the teacher in the course structure. Similarly, the other two charts show the selected section support level before and after the modifications.

Figure 1.4 illustrates the visualization part of the tool interface for a selected section in General Mode. Each chart consists of two parts. The upper part of a chart consists of a set of bars to show the strength of the harmony of the course/unit/section with each of the eight learning style poles (i.e., active, reflective, visual, verbal, sensing, intuitive, sequential, and global), in terms of percentage (calculated by the average of the three factors illustrated in Sect. 1.3.1). Each learning style dimension is represented by two horizontal bars, one for each pole, where the two poles show the two different preferences of the dimension. The longer the bar, the more the course/unit/section fits with the learning style. The lower part of a chart contains only one bar that shows the overall support level of the course/unit/section for diverse learning styles (calculated by the average of the support level of the eight poles). Once the teacher moves the cursor over any bar, a tooltip appears to display more details about the analysis factors illustrated in Sect. 1.3.1.

Assume that the teacher made a modification in the selected section by adding an "Animation" LO to that section. The existence of an "Animation" LO can support active, sensing, and visual learning styles (Felder and Silverman 1988). After reanalyzing the course considering the teacher's modification and by comparing the two charts in Fig. 1.4, it can be noticed that the section support levels for active, sensing, and visual learning styles were improved (as shown in the right chart by the black arrows). Consequently, the overall section support level for diverse learning styles improved as well (as shown at the bottom of the right chart).

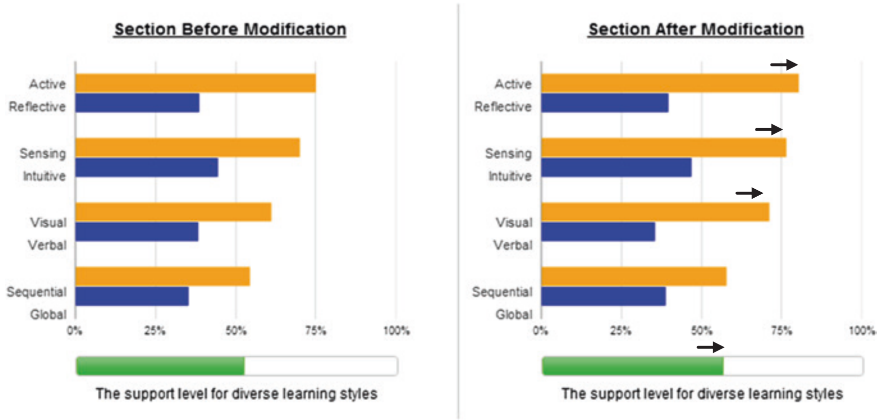

Fig. 1.4 Visualization part of a selected section (general mode)

The teacher can switch to Cohort Mode to visualize the course support level for enrolled students' learning styles. Figure 1.5 illustrates a screenshot of the visualization part of a selected section in Cohort Mode. The charts visualize the data about students' learning styles in comparison with the course support level (calculated by the average of the three factors illustrated in Sect. 1.3.1). As shows in Fig. 1.5, each learning style dimension in a chart contains two bars; the upper one shows the course/section support level for each poles of that dimension (for examples, "active" on the right and "reflective" on the left); the lower bar shows the learning styles of the respective cohort of students, in terms of different levels

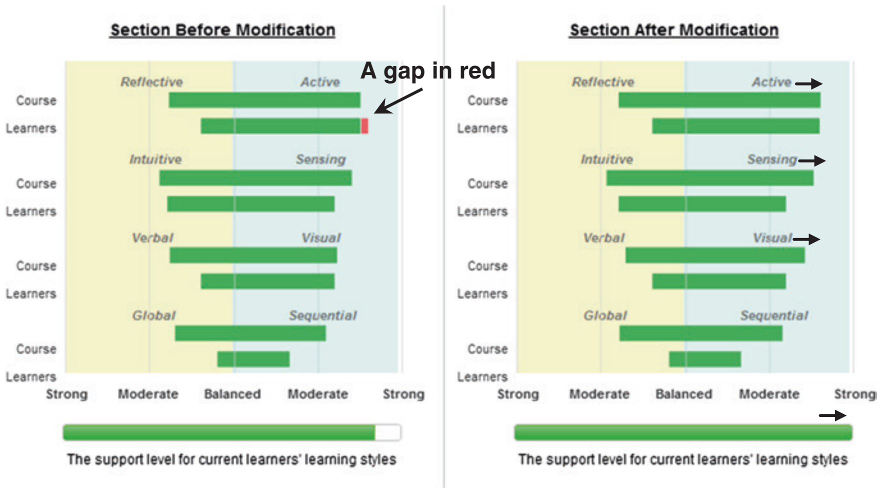

Fig. 1.5 Visualization part of a selected section (cohort mode)

varying from strong to balanced. In case that all students are fully supported, the bar is displayed in green color, otherwise a gap is shown in red. The intensity of the red color indicates the number of unsupported students.

For example, the chart at the left side (Fig. 1.5) shows that reflective, sensing, intuitive, visual, verbal, global, and sequential learners are well supported by the course. On the other hand, active learners are not fully supported; there is a gap between the course support level for the active learning style and the learning style of the respective cohort of students. Once the teacher moves the cursor over any bar, a tooltip appears to display more information about the level of support and the number of supported and unsupported students. Considering the example mentioned in the previous section and the teacher's modification in that example, the chart at the right side (Fig. 1.5) shows that the gap between the course support level and the learning styles of the respective cohort of students was eliminated, and the students were fully supported (as shown at the bottom of the right chart).

The visualization interfaces make teaches aware of how their modifications would impact the course support level for students with different learning styles. Once the required support level is reached, the teachers can navigate to the course webpage in the LMS and actually implement the necessary modifications in the course structure to improve the course support level for better supporting students with different learning styles.

1.4 Automatic Detection and Visualization of Device Functionalities and Usage

While the previous two sections consider analyzing learners' characteristics and courses in LMSs based on learners' cognitive abilities and learning styles, this section focuses on the context of a learner. The context plays a relevant role in ubiquitous and mobile learning environments. The recent advances in mobile technologies have allowed the widespread use of mobile devices around the world. Accordingly, learning can take place anytime and anywhere using mobile devices (e.g., smart phones, tablets) to facilitate human interaction and access to learning contents with fewer restrictions of time and location (Chen et al. 2008).

Mobile learning, as it is known, is a crucial aspect of ubiquitous learning. Such mobile settings bring important improvements for ubiquitous learning by providing a more flexible and authentic experience for the learner. However, additional factors (such as the learner's context) influence the learning process in ubiquitous learning environments and should be considered, especially when aiming at considering adaptation and personalization aspects. We address this issue using data collected from the usage of the learners' mobile devices. The usage of the learner's devices provides important information about the user and his/her context (Ogata et al. 2010). The functionalities (features) of a device work as sensors gathering different data such as location, ambient light, humidity, temperature, and connectivity. Such data can be used to build a context profile (Roman and Campbell 2002).

Details about the learner context help in the process of personalizing the learning experience. Furthermore, the way the learner uses the device—its applications and features—is a good indicator about his/her preferences and other personal traits (information to build a user profile). The user profile and the context profile are elements which compound a comprehensive model about the learner and his/her context that can be considered as the basis of a personalized adaptive learning system (El-Bishouty et al. 2007; Graf and Tortorella 2012).

Several applications are developed in order to monitor the usage of smart mobile devices aiming at enchanting the device usage. Due to the advantages of open source, many of them are implemented for Android-powered mobile devices (Joiku Phone Usage 2013; Elixir 2012). In contrast to such applications, the proposed framework in this section illustrates a methodology to automatically detect the existence of types of device features and functionalities that the learner may use in his/her device for learning. It is not targeting a particular feature, or a certain device type or operation system. In addition, it monitors how the learner interacts with and uses each feature in order to understand and recognize the learner's preferences on how to use mobile devices for learning. Such context information opens up new possibilities for providing adaptive learning environment.

1.4.1 Framework Architecture

The proposed framework (Fig. 1.6) is a client–server generic framework designed to run on smart phones, tablets, laptops, and desktop computers (Lima et al. 2013). On the client side—running on the learners' devices—the application is divided into two parts. The first one is the User Part, where personal information about the user (i.e., learner) is gathered and usage data visualization is presented. This part has two components: the Personal Information Manager, allowing the learner to login and to provide personal information (e.g., login name, full name, email address, and other characteristics). This information is grouped under the name of User Object. The second component is the Visualizer, which is responsible for providing a user interface to show the feature availability and usage information.

The second part of the application is the Device Part, where device information about the availability and usage of features is collected. This is done by Feature Detectors and the Tracker. Each Feature Detector is associated with a feature (e.g., internet connection types, keyboard presence, and available touch screen) and able to discover whether the respective feature is available in that device. The data collection from all Feature Detectors is used to create the Device Object, which includes information to describe the tracked device in terms of its features. Furthermore, the Feature Detectors provide information about the usage of the respective features to the Tracker.

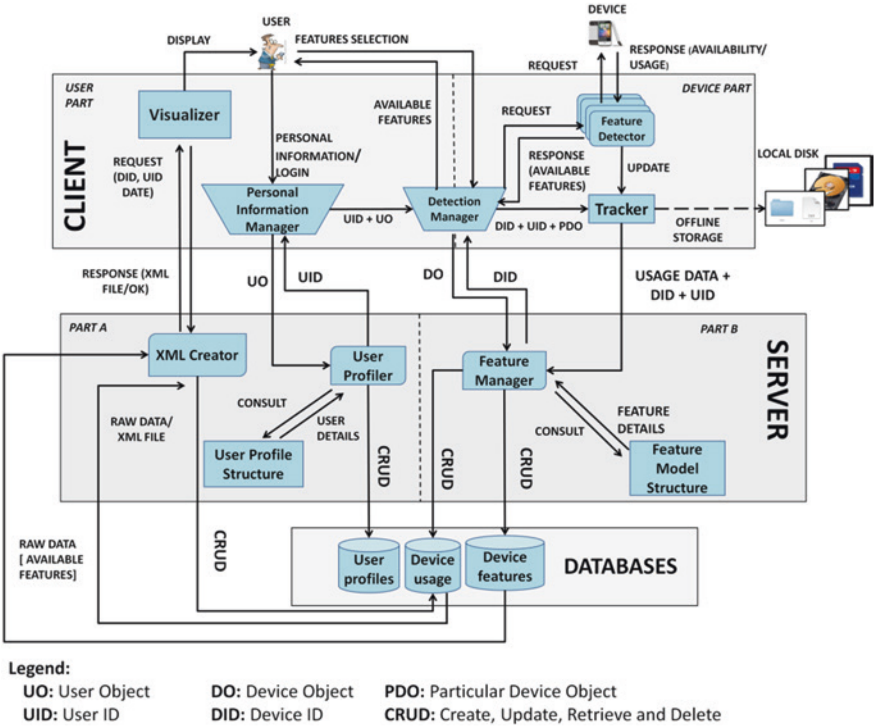

Fig. 1.6 The framework architecture

The Tracker gathers the learner's device usage data by receiving updates from the Feature Detectors and stores this information on the server. However, if there is no connection to the server available, the Tracker creates a local backup file with the gathered data and postpones the online storage until connection to the server is available. Moreover, there is a component which belongs to both parts of the client side, the Detection Manager. The Detection Manager allows the learner to select what features the system should track. In order to do so, the Detection Manager detects what features are available in the device by calling the Feature Detectors when the learner starts the application for the first time and displays these features for selection. The set of information about the features selected by the learner is grouped under the name of Particular Device Object and represents a subset of the information in the Device Object. Furthermore, the Detection Manager encapsulates the user id, the device id, and the Particular Device Object and sends all this information to the Tracker.

The server side mainly manages the database and the identification of devices and users, and is divided into two parts. Part A communicates with the User Part of the client and consists of the following components: the User Profiler, which is responsible for receiving the user information as a User Object from the client and for storing it in the database. The second component is the User Profile Structure, which stores what personal information of the learner is considered in the framework; and the third component is the XML Creator which processes the availability and usage data and generates a XML file which is used by the Visualizer.

Part B of the server side is responsible for the communication with the Device Part of the client. It is composed of the Feature Manager and the Feature Model Structure. The former is responsible for receiving the device information in form of a Device Object, including the available features on the device and its usage data from the client, and for storing this information in the database. The Feature Model Structure describes the device features considered in the system. Moreover, there is a database, divided into three parts: First, it stores the device information, including the device id and which features are available to be used. In other words, it stores a set of Device Objects. Second, it stores user information (e.g., login name, full name, email address, and other characteristics), in other words, a set of User Objects. Third, it combines the data relating a device to a learner and stores the usage data of the device by the learner.

1.4.2 Application for Android Phones

As mentioned above, the proposed framework is designed for tracking and analyzing the availability and usage of device features from different devices such as mobile phones, tablets, and desktop computers. In this section, we describe an application for mobile phones, running on the Android operation system.

This application—the Usage Observer—aims at providing the possibility for learners to select features on their devices, monitor the usage of those features, and view data about feature availability and usage in a user-friendly way. Moreover, learners can allow teachers to access these data. The application is composed of seven interfaces: login, main, feature selection, personal information/preferences, tracker management, feature usage visualization, and administration. Some of the major interfaces can be seen in Figs. 1.7 and 1.8.

The login and registration process is managed by the Personal Information Manager and the Detection Manager. If the learner does not have an account yet, he/she needs to go through the registration process. In the first step, the learner can provide personal information and preferences (as shown in Fig. 1.7b), leading to the creation of a User Object by the Personal Information Manager. This interface (as well as the other user interfaces) is implemented as an Android component extension called Activity.

In the second step of the registration process, the learner registers his/her device by accessing the Detection Manager, where he/she can select which features should be tracked (Fig. 1.8a). This interface presents the device's available features, identified by the Feature Detectors, and allows the learner to select which features he/she wants to be tracked. As a result, a Particular Device Object is created containing information on those selected features. The Feature Detectors

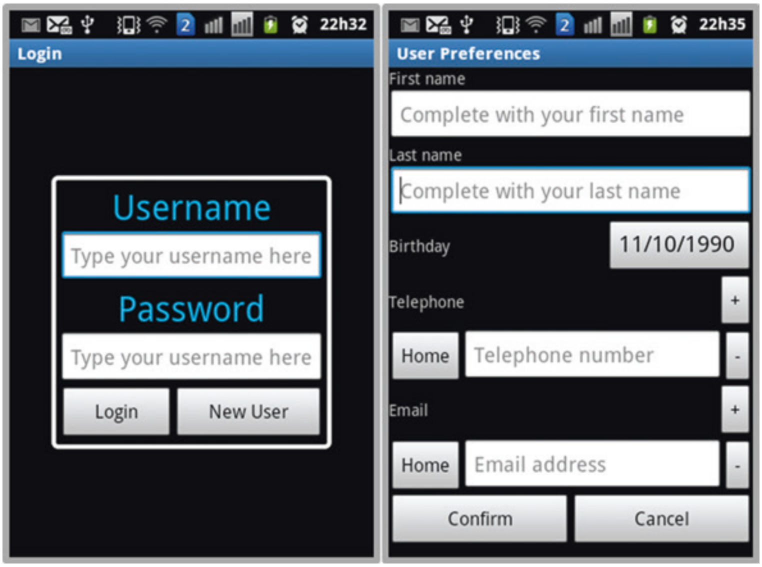

Fig. 1.7 Application interfaces—**a** login interface; **b** personal information interface

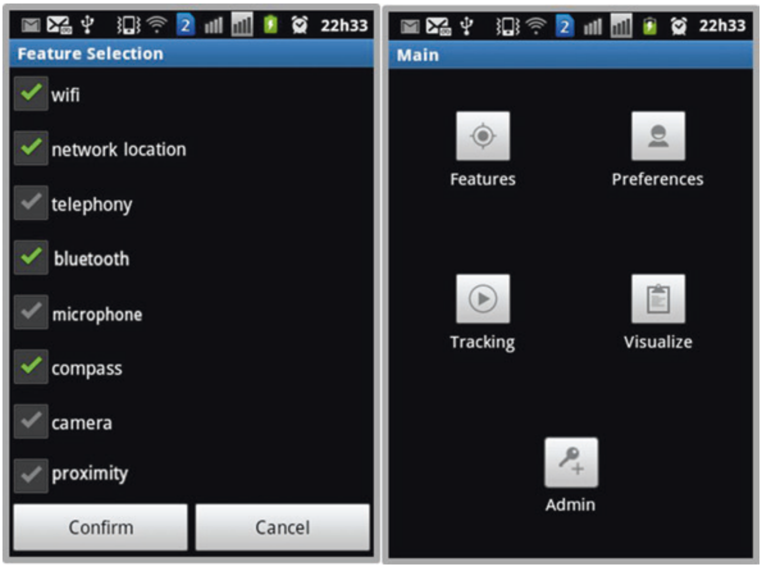

Fig. 1.8 Application interfaces—**a** Feature selection interface; **b** Main interface (access point of the other interfaces)

are implemented as one single service, called Detection Service. Services are an Android component extension, which run in the background and typically perform long-term tasks. If a learner is using different devices, the registration for the user account has to be done only once but whenever the learner using a new device, this device has to be registered by performing the second step of the registration process. Once the learner and his/her device are registered, he/she can login by providing the correct username and password (Fig. 1.7a) and is presented with the main interface (Fig. 1.8b), where he/she can navigate through the application.

The main interface allows the learner to access the personal information entered in the registration process (through the Preferences symbol) and the feature selection interface (through the Features symbol) in order to change personal information and selected features. Furthermore, the main interface provides access to the tracking and visualization interfaces. The tracking interface presents raw data about the tracked features (accessible through the Tracking symbol) and the visualization interface shows user-friendly visualizations about the availability and usage of features on the learner's devices (accessible through the Visualize symbol). Administrators and teachers can additionally access the administrator interface (through the Admin symbol), which provides them with additional visualization interfaces to compare usage information from different learners and/or different devices.

While learners are using their devices, the system tracks the usage of the selected features. This is done by the Tracker Service, which represents part of the Feature Detectors and part of the Tracker component in the framework. The Tracking Service is responsible for collecting data about the usage of the previously selected features. One of the biggest development challenges was to implement the algorithm for the Tracking Service to track the usage of the selected features. The Android platform does not provide any direct way of collecting this type of data for certain features. After evaluating several different approaches, we implemented this service by reading the system generated logs and gathering the timestamps related to the use of each selected feature. Hence, the Tracker component in the framework works as a logger which means the developed Tracker service collects log data provided by the system and sends it to the framework's server side. The server is then responsible for parsing the received log data and for obtaining the selected feature's usage data and statistics based on the timestamps contained in the log data. This task is performed by the Feature Manager component, which then stores the new information in the database.

1.5 Conclusions

This chapter illustrated three different approaches to automatically analyze learners' characteristics and courses, building a foundation for providing learners with adaptive and personalized support based on their cognitive abilities, learning styles, and context.

A framework is proposed for identifying students' WMC from their activity log information in learning systems, as well as its implementation into the detection tool DeLeS. Six behavior patterns have been identified to be, on one hand, relevant for the identification of WMC as concluded by the literature, and on the other hand, to be domain and learning system independent so that our proposed approach is generic and can be used in different learning systems. By knowing students' WMC in learning systems, teachers can individually support students and provide them with personalized recommendations, while students can better understand their weaknesses and strengths, and use this information to improve their learning. Furthermore, information about students' WMC can be used as input for an adaptive learning system to automatically provide students with individualized materials and activities as well as personalized recommendations, considering their WMC.

With respect to learning styles, a mechanism and an interactive tool are presented for analyzing existing course contents in LMSs based on students' learning styles. The tool provides teachers with an interactive graphical user interface that allows them to play around with the course structure. This interactive interface supplies the teachers with drag and drop utility to add, move, and/or remove LOs, and then visualizes the changes in the course support level. This supports the teachers to decide which modifications should be implemented in the actual course structure. The aim of these modifications was to meet the need of different students' learning styles, which enables the teachers to provide personalized support to meet the need of students with different learning styles.

For the identification of context, a framework and an application are presented for identifying learners' preferences on how to use their devices for learning. This approach aims at building a comprehensive context profile through detecting available features of a device (e.g., keyboard, touch screen, internet connection, camera, GPS, and so on) and tracking the usage of these features by its users. Furthermore, the application visualizes the gathered data in a user-friendly way. The gathered information is the basis for extending learning systems with advanced adaptive and intelligent capabilities that allow personalizing user interfaces and providing learners with adaptive course structures and recommendations based on availability and previous usage of device features.

Acknowledgments The authors acknowledge the support of nserc, icore, xerox, mitacs, and the research-related gift funding by mr. a. markin.

References

Atman, N., Inceoğlu, M. M. & Aslan, B. G. (2009). Learning styles diagnosis based on learner behaviors in web based learning. *Proceedings of ICCSA 2009, 5593/2009, 900–909*. Springer, Heidelberg.

Broadway, J. M., & Engle, R. W. (2011). Lapsed attention to elapsed time? Individual differences in working-memory capacity and temporal reproduction. *Acta Psychologica, 137*(1), 115–126.

Chen, G. D., Chang, C. K., & Wang, C. Y. (2008). Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques. *Computers and Education, 50*(1), 77–90.

Clay, J. & Orwig, C. J. (1999). Your learning style and language learning. *Lingual Links Library, Summer Institute of Linguistic, Inc (SIL) International version 3.5*.

El-Bishouty, M. M., Ogata, H., & Yano, Y. (2007). PERKAM: Personalized knowledge awareness map for computer supported ubiquitous learning. *Educational Technology and Society, 10*(3), 122–134.

El-Bishouty, M. M., Chang, T.-W., Kinshuk, & Graf, S. (2012). A framework for analyzing course contents in learning management systems with respect to learning styles. In G. Biswas et al. (eds.) *The 20th International Conference on Computers in Education* (ICCE 2012), pp. 91–95. Asia-Pacific Society for Computers in Education, Singapore.

El-Bishouty, M. M., Saito, K., Chang, T.-W., Kinshuk, & Graf, S. (2013). An interactive course analyzer for improving learning styles support level. In *Proceedings of the 3rd International Workshop On Human-Computer Interaction And Knowledge Discovery In Complex, Unstructured, Big Data* (HCI-KDD 2013), Lecture Notes in Computer Science, Springer, Vol *7947*, pp. 136–147.

Elixir (2012) https://play.google.com/store/apps/details?id=bt.android.elixir. (Accessed on Dec. 2013).

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology, 51*(1), Working Memory: Beyond Language and Symbolism, 17–26.

Felder, R., & Silverman, L. (1988). Learning and teaching styles. *Journal of Engineering Education, 94*(1), 674–681.

Felder, R. M., & Soloman, B. A. (1997). Index of learning styles questionnaire. North Carolina State University, http://www.engr.ncsu.edu/learningstyles/ilsweb.html. (Accessed on October, 2013).

Gathercole, S. E., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. London: Sage Press.

Graf, S., & Kinshuk, K. (2007). Providing adaptive courses in learning management systems with respect to learning styles. In: G. Richards (Ed.), *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (e-Learn), 2576–2583, AACE Press, Chesapeake, VA.

Graf, S., Kinshuk, & Liu, T.-C. (2009a). Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach. *Educational Technology and Society*, *12* (4), 3–14.

Graf, S., & Tortorella, R. (2012). *Personalized mobile learning via an adaptive engine*. Exploring the future of Technology Enhanced Learning ICALT: IEEE Computer Society.

Graf, S., Lin, T., Kinshuk, Chen, N. S., & Yang, S. J. H. (2009b). Learning styles and cognitive traits—their relationship and its benefits in web-based educational systems. *Computers in Human Behavior*, *25*(6), 1280–1289.

Huai, H. (2000). *Cognitive style and memory capacity: Effects of concept mapping as a learning method*. Doctoral Dissertation, Oct 2000, Twente University, The Netherlands.

Joiku Phone Usage (2013) https://play.google.com/store/apps/details?id=com.jupiterapps.Phoneusage&hl=en. (Accessed on Dec 2013).

Kuljis, J., & Liu, F. (2005). A comparison of learning style theories on the suitability for eLearning. In M. H. Hamza (Ed.), *Proceedings of IASTED*, pp. 191–197. ACTA Press.

Lima, R. H. P., El-Bishouty, M. M, & Graf, S. (2013). A framework for automatic identification and visualization of mobile device functionalities and usage. *Proceedings of SouthCHI 2013*, Lecture Notes in Computer Science, Springer, 7947, 148–159.

Lin, T., Kinshuk, & Patel A. (2003). Cognitive trait model—a supplement to performance based student models. In *Proceedings of international conference on computers in education* (pp. 629–632). Hong Kong.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers and Education, 54*, 588–599.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Ogata, H., Li, M., Hou, B., Uosaki, N., El-Bishouty, M. M., & Yano, Y. (2010). SCROLL: Supporting to share and reuse ubiquitous learning log in the context of language learning. *Proceedings of mLearn, 2010*, 40–47.

Onyejegbu, L. N. & Asor, V. E. (2011). An efficient model for detecting LEARNING STYLE preferences in a personalized E-Learning management system. *Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Software Engineering* (JSSE), May Edition.

Paredes, P., & Rodríguez, P. (2004). A mixed approach to modelling learning styles in adaptive educational hypermedia. *Advanced Technology for Learning, 1*(4), 210–215.

Park, O., & Lee, H. (2003). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 651–684). Bloomington, Indiana: AECT.

Roman, M. & Campbell, R. H. (2002). A user-centric, resource-aware, context-sensitive, multi-device application framework for ubiquitous computing environments. CS Technical Report.

Tseng, J. C. R., Chu, H. C., Hwang, G. J., & Tsai, C. C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers and Education., 51*, 776–786.

Unsworth, N., Redick, T. S., Gregory, J., Spillers, J., & Brewer, G. A. (2012). Variation in working memory capacity and cognitive control: Goal maintenance and microadjustments of control. *The Quarterly Journal of Experimental Psychology, 65*(2), 326–355.

Woehrle, J. L., & Magliano, J. P. (2012). Time flies faster if a person has a high working-memory capacity. *Acta Psychologica, 139*(2), 314–319.

Chapter 2
An Approach to Measure Coding Competency Evolution

Toward Learning Analytics

Vive Kumar, Kinshuk, Thamaraiselvi Somasundaram, Steve Harris, David Boulanger, Jeremie Seanosky, Geetha Paulmani and Karthikeyan Panneerselvam

Abstract There is a great deal of interest in the area of learning analytics and their use in assessing both student and content success in a digital learning environment. This chapter describes the results of a pilot study that used a combination of software tools and processes to collect the data generated from students taking an introductory C programming course, and their interactions related to specific study activities. This study helps determine whether it is possible to collect enough useful analytics to begin to identify constructive learning practices and strategies that determine competency evolution. We look at the types of learning traces collected through our system and then discuss how the data might be used to provide insight into student, class, or content actions. We then consider how additional analytics may be collected and used to supplement our initial results. The technologies used in this study address requirements of big data learning analytics where the data come from (a) the learner and his/her immediate surroundings; (b) the social network of the learner related to the learning tasks; and (c) the environment that inherently encases learning activities. While the data are used to measure competencies and their evolution over a period of time, the resultant profiles show the evolutionary processes students have adopted in developing specific competencies.

Keywords Learning analytics · Big data · Coding · Novice programmer

V. Kumar (✉) · Kinshuk · S. Harris · D. Boulanger · J. Seanosky · G. Paulmani
School of Computing and Information Systems, Athabasca University, Athabasca, Canada
e-mail: vive@athabascau.ca

T. Somasundaram · K. Panneerselvam
Madras Institute of Technology, Anna University, Chennai, Tamil Nadu, India

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_2

2.1 Introduction

Students consume an array of media in their studies, whether in classroom or in online instruction. The media includes slide shows, reading material, audio/video lectures, interactive tutors, and automated tests. As they work through the course material, they generate digital learning activity traces that track their interaction with the course material, much like a Web analytics system tracks visitors to a Web site. Similarly, as students complete the associated computer programming assignments, software development traces are generated which can be used to track the student success and coding competency.

Software development traces generated by students as they solve computer programming problems, such as number of compiles, warnings, and various errors, can be captured from their development environments, along with the student learning traces from the digital course materials accessed for the duration of the course.

In this pilot study, a software analytics system called HackyStat[1]—an open source project originally developed by researchers in the University of Hawaii—collects software development activities and logs from a variety of development environments, including Eclipse, EMACS, and Visual Studio. In addition, an open source Moodle plug-in developed by the University of Las Palmas de Gran Canaria called Virtual Programming Lab (VPL)[2] was altered to perform a similar data collection task from within Moodle itself, for less intensive programming exercises. Moodle's reporting module itself offers data traces on student interactions with the learning material housed in that system as well as performance measures of learners.

Using a combination of static analysis tools and toolkits, depending on the language in use (e.g., Java program analysis uses a combination of Checkstyle, FindBugs, and JUnit, while C programs use a combination of Lint and Virtual Programming Lab (VPL) customizations), coding activities of students can be logged at finer levels of granularity. Coding activities include code design, writing of code, efforts of debugging, attempts at documenting the code, and testing of the code. Using both manual and automatic grading, the work of the students can also be graded at finer levels of granularity. All these datasets can be captured in HackyStat.

The HackyStat system allows for the greatest flexibility in data collection and retains the raw data in a SQL database for further processing. The collected data include not only the final product, a software program, but also the intermediate steps, such as designing, coding, debugging, testing, and documenting, that the student had taken to reach the final product. Also, because of its open source nature, additional HackyStat plug-ins can be developed or altered to track programming in other environments or languages, as required.

[1] http://code.google.com/p/hackystat/.

[2] http://vpl.dis.ulpgc.es/.

The raw datasets are used to populate a coding competency ontology (based on Java and C ontologies[3] from the University of Pittsburgh). The ontology contains two parts: first, the establishment of a hierarchical framework for the domain knowledge that primarily includes classes of programming actions, traces from the learning materials, types of programming problems, and primary relations between these categories and, second, a causal mechanism that attempts to develop models of causation in coding-related activities.

2.2 Research Rationale

The goal of this research is to make inferences about student's coding competency that may establish causal relationships between student activities and learning outcomes and to identify and share inferences that seem to result in constructive coding practices for students. Additionally, we are also interested in the identification of subgroups of students for whom different learning strategies lead to optimum coding competency or that certain types of materials are more effective learning tools than others. Further, suggestions could be proposed to the student, perhaps through the use of a smart agent, to focus on certain materials and strategies based on their performance and that of past students. We reviewed literature to have an understanding of the technological and conceptual hurdles that need to be tackled in achieving these goals. The following summarizes our review efforts.

Whether observed or not, data are being continually produced in all phases of learning whether it is classroom-based lecture or an online discussion or a blended training exercise or lifelong work scenarios. Instructional design efforts generate their own sets of data concerning course design models, development workflow, course guidance, adaptation opportunities, course comprehensiveness, and course quality. Irrespective of where, why, how, and when learning happens, learning environments do create new learning experiences and correspondingly new sets of data.

Each learning experience can be measured as long as the underlying data are collected. Be it personalized learning, formal learning, ad hoc learning, classroom learning, or mobile learning, different types of learning environments offer different types of learning and instructional attainment datasets. The associations that evolve among these datasets on a day-to-day basis have conveniently and predominantly been ignored since the beginning of modern era. For instance, from a learner's perspective, summative assessments, augmented with formative assessments, are used to only estimate competency of the learner. The capacity of the learner and the effectiveness of learning processes undertaken by the learner are not normally measured. This is not because of the challenges involved in obtaining and validating data on capacity and effectiveness, but because of the pervasive nature of 'lethargy' in teaching.

[3] http://www.sis.pitt.edu/~paws/ont/java.owl and http://www.sis.pitt.edu/~paws/ont/c_programming.rdfs.

Learner competence can be measured to a greater extent by using a variety of assessments. Learning capacity, on the other hand, depends on a number of personal factors including working memory, motivation, support, learning material, and learning styles. Learning effectiveness is a measure that mainly involves competence and capacity. Capacity and effectiveness have generally been ignored by contemporary instruction attributing to the lethargy mentioned above. Expectation of the ‘bell curve’ as the norm of classroom performance is another factor that attributes to lethargy in teaching. Constraints placed on learning outcomes force instructional processes to be satisfied mostly with summative feedback-based associations, offering further attributions to lethargy.

Learning processes/activities undertaken by individual learners from the time a concept to be learned is introduced, up to the time instructional efforts cease to teach that concept to that learner can potentially be observed using a variety of techniques and technologies. For instance, the behavior of classroom interactions of students attending a lecture can be video-recorded and analyzed at real time to estimate the level of comprehension of a set of concepts introduced in the lecture. The behavior of online interaction of students in a discussion forum can be analyzed at real time to estimate their writing competence. The behavior of social interactions of students can also be analyzed to estimate their sentiments on specific services offered by the institution related to learning. These exclusive observations, along with other traditional data, can supply the ‘raw data’ for learning analytics.

Raw data are being produced at an alarming rate in other domains including transportation, health care, geographic advertising, retail, banking, and energy (Manyika et al. 2011). A report on online learning in Canada[4] outlines the market share of learning management systems (LMS) that already produces large quantities of data on learning and the inclusion of online learning as part of core business plan of academic institutions across Canada.

The following sections review techniques, technologies, and systems that specifically address raw data collection with a view to big data analytics.

2.2.1 Techniques, Technologies, and Systems

An infrastructure (Bienkowski et al. 2012) that supports learning resource discovery, sharing, and amplification indicates to a large volume of metadata and data that are not usually barricaded from public view. At-risk students can be recognized based on data about the type of risk, interventions based on predictive models can be designed to mitigate that risk, and the utility of the models can be gauged by the trace data on applied intervention (Essa and Ayad 2012). A number of data-driven analytics drivers to mine effectiveness of learning have been reported for use in learning and academic analytics (Ferguson 2012).

[4] http://www.contactnorth.ca/online-learning-canada.

Social learning analytics offers access to a slew of data on informal conversations among students and instructors (Shum and Ferguson 2012).

Raw data are not confined only to formal and informal datasets, but also include ad hoc datasets with marginal overlap with learning. Ad hoc datasets include browsing patterns, reading habits, writing style (including freehand writing), coding, posture analyses, collaboration drivers, thinking protocols, chats, domain-specific tool traces, recording of cognitive and metacognitive traces, and knowledge traces.

Among many others, raw data can be processed toward data mining (Romero et al. 2008), program evaluation (Hung et al. 2012), real-time classroom feedback (Wood et al. 2013), prediction (Watson et al. 2013; Abdous et al. 2013), visualization (Mazza and Dimitrova 2003; Kim and Lee 2013), and modeling (Medeiros et al. 2013), among others.

Potential observations and analyses of the aforementioned categories would shed light on instructional processes that indicate introduction, deliberation, assimilation, evaluation, and application of a concept or skill to various degrees by different types of learners. However, a holistic view of the rates of changes in adaptations, repetitions, and refinements of instructional processes and instructional resources, as well as the rates of changes in learning processes and learner capabilities, has never been observed as a matter of fact—till now, till the arrival of learning analytics as the science of analysis that concerns learning, and the suite of tools that observe, validate, and semi-automatically associate datasets related to learning and instruction.

Presently, it is not possible to technically trace the evolution of individual skills of programmers through the accumulated study experiences of individual or groups of students, thus necessitating the need for a more effective capture of study experiences of learners. In particular, we plan to use ontologies (Dicheva et al. 2009) since ontologies facilitate the integration of various sources of information covering the same domain knowledge. In our case, we use ontologies to interrelate information about learning objects, learning activities, and learners, captured from various tools embedded within Moodle.

While these technologies focussed mostly about processing readily available data about study experiences of learners, quite a significantly large volume of data about formative study experiences of learners, particularly online learners, goes unnoticed, unutilized, and unappreciated. A number of domain-specific tools target and enhance activities associated with learning (Zinn and Scheuer 2006; Mazza and Milani 2005; Mazza and Dimitrova 2004; Kosba et al. 2005; Brooks et al. 2004; Jovanovic et al. 2009; Shakya 2005; Baghaei et al. 2007; Ghidini et al. 2007).

Competency is an elusive term. There is plenty of research on how to computationally capture it, process it, and foster it (Butler et al. 2008; Boekaerts and Corno 2005; Cross and Angelo 1988; Angelo and Cross 1993). Yet, there is a considerable gap in how computational models capture programming competencies. The aim of this research is to tackle this and develop a widely acceptable computational model for Java competency that can be customized by the individual student.

Study of programming skills is not new. The most traditional method of evaluating student's competency is that of standard examination by way of assignments, projects, laboratories, examinations, and tests. In 2004, researchers Rountree, Rountree, Robins and Hannah (Rountree et al. 2004) utilized this standard method of evaluation to investigate the comparison of difficulty between code generation and code comprehension for introductory Java programming course students. They soon found that this method of evaluation showed 'Contrary to their initial expectations, the code comprehension and generation skills appear to be tracking each other', as well as 'students receive a grade for their work, but this result tells us little about the nature of their knowledge'.

Rao et al. track the growth of competencies by observing how students develop multiple computer programs, augmented with summative as well as formative feedback (Rao et al. 2007). Mixed-initiative interaction (Allen et al. 1999), employed by Rao et al., presents researchers with opportunities to intervene and prompt learners based on certain criteria, where the intervention is initiated by the system. Java learning environments such as MICE and CORE bring into view excellent mediums by which students' development activities can be tracked (Rao et al. 2007).

Novice java programming students write numerous blocks of code and submit them to the compiler for verification where in return the compiler provides them with a summative feedback on the correctness of their code based on a set of accepted complex syntaxial style guides (Rao et al. 2007). These articles and their conclusions showcase that programming competency could be tracked and prompted to ensure that students are aware of their current competency levels, aware of what they need to do to improve the levels, and aware of what tools and skills they need to engage to target such an achievement.

As education moves out of the classroom and enters the era of online and distance education, the type of personalized and adaptive scaffolding needed to support different needs of a variety of students is an area of research that is larger than life. This research will contribute to the study of personalized interactions, customizable interfaces, and adaptive instructions as key premises. In summary, this research will study how continuous data collection and modeling of the collected data in online activities related to coding can benefit task-specific competencies for novice programming students.

2.3 What Is Big Data Learning Analytics

Learning analytics is the science of analysis, discovery, and utilization of learning traces in emergent and related levels of granularity. Analysis could include techniques ranging from data mining to machine learning and to big data analysis. The discovery of new relations and the discovery of even new data include unconventional data, for example, the family income of a politically competing region, inherent economic drivers influenced by a curriculum, and rate of changes in motivation levels of students with respect to weather. Relations of interest include sentiments among learners across collaborating groups, interinstitutional credit transfer policies among institutions, and mutual respect among instructors. Trace data refer to observable raw data of study activities such as reading, writing, conceptualizing, critically thinking, solving problems, storytelling, and visualizing, where a network of study activities leads to a measurable chunk of learning. For instance, the types of sentence openers used by a learner, the range of errors that the student can confidently correct, the level of trust exhibited by the student in sharing information in a forum, and the depth of understanding in a set of concepts are examples of learning traces where one could measure learning over time. In learning analytics, data are expected to arrive continuously, potentially in an interleaved fashion, subject to interpretation at various levels of granularity.

In general, learning traces translate raw data into incoming data for learning analytics, where incoming data are typically big, unstructured, unrelated, and fit multiple models and possibly multiple theories. Importantly, learning traces capture highly personalized study experiences. For instance, consider the example of studying about the notion of a pointed object penetrating better than a blunt object. The study goal is to understand why this is so. A visual-oriented learner may choose to study and explain this phenomenon using visual tools, while a fellow student may work out the mathematics behind pressure and explain the results in terms of equations. Thus, learning traces capture different types of skills and background knowledge exhibited by individual learners, the extent to which learners learned new concepts while studying a particular material, the amount of time (efficiency) learners took to study the material, the effectiveness of assessments to conclude that the concept had been learned efficiently, evidences of learning activities/experiences (e.g., a video of the learner trying a sharp pencil on someone's palm and the learner's answer to the question on pressure in the examination), and resources used by learner (e.g., the pencil, simulation run by the learner). While each learning trace is measurable, there is no standard scale of measurement that applies across different types of traces. A learning trace is the least common denominator for a measure of learning.

Data mining offers techniques that typically operate on well-defined, medium-sized datasets (up to gigabytes). Intelligent tutoring systems typically operate on limited-sized data (up to megabytes). Learning analytics aims to operate on large volumes of data (at least in the order of terabytes, hence the name, big data) as well as large volumes of models produced from the data. The models are dynamic, if not emergent, because more often than not, processing of data would include newly arriving, marginally related datasets. Examples of emergent models include 'reading habits across domains', 'writing styles across contexts', 'application of problem-solving skills across tools' (SPSS, R, Eclipse, MATLAB, and so on), 'social distances while collaborating with peers', 'comfort zone when interacting with instructors', and 'research potential'.

Data can be gleaned with sensors. Contemporary technologies only offer sensors that observe a tiny fraction of data associated with these new learning experiences. In learning analytics, one could conceive of three types of sensors—instructional sensors, context sensors, and internal sensors. Instructional sensors have explicit association with the study as seen in LMSs. Intelligent tutoring offers a wide variety of instructional sensors such as sensors to observe self-regulation activities of learners and sensors that observe the interactions between learners and anthropomorphic pedagogical agents. Context sensors include the observable environment in which learning takes place. Examples of context sensors include autonomous software agents that act independently on students' behalf and data traces that cluster students from across multiple institutions, based on specific competencies. Internal sensors include methods that estimate learners' ability to think laterally and critically, ability to strategize in a game environment or regulation exercise, ability to use working memory and comprehension techniques, and the ability to pace oneself and proactively study. The observations of these three types of sensors and the interrelations among individual data streams produced by each sensor yield the raw data.

While big data have been a recognized field of exploration in many domains such as traffic regulation, health care, geographic advertising, retail, banking, public sector, consultancy, energy efficiency, and policing, applications of big data techniques in learning are still in their infancy. Learning attainment and instructional attainment need to synchronize with each other to offer an optimal approach to learning. In addressing this very goal, this chapter highlights the application of learning analytics with a case study. The case illustrates the collection of raw data, the transforming the raw data into learning traces, the application of learning analytics techniques on learning traces, the use of results of analytics toward balancing learning and instructional attainment, and the measurement of impact of analytics.

Technology-enhanced learning, in general, implies that learners study learning content and collaborate with their peers and instructors through the use of learning management systems. As part of their study, learners need to solve problems by using development environments built for specific programming languages they study. Learners also often need to use resources publicly available on the Web (not necessarily included into the official course materials) and ask for the help of their peers. However, with the current technology support, students are not able to seamlessly integrate their learning activities when they use different tools. Further, the current technology does not allow formal integration of interaction data collected from across these tools. Similarly, for their instructors, it is very hard to relate the underlying activities their students were engaged in that produced the submitted learning products. This is not conducive for instructors to be able to help students with personalized and context-specific advice. In general, most instructors only receive the final product of students' study efforts; neither the study process students went through nor the intermediate study products and study issues faced by students come to the attention of instructors.

2.4 Datasets and Analysis

We intend first to describe the overall data flow and then discuss about the various technologies employed in the processing of the data all along the way that is the chain of I/O. We first note that a learning analytics system not only collects data to support the mission of the owner organization but also collects metadata about the interactions that the various stakeholders have with any component of the system. Although the learning analytics framework has been designed to be as generic as possible, we discuss its use in a programming context which makes up a perfect, complex experimental environment to show its potential.

A number of HackyStat sensors have been developed that include Eclipse Integrated Development Environment (IDE) sensor called EIDEE, a Java tutor sensor called MILA, and Moodle-embedded IDE sensor called VPL. In addition, Moodle itself minimally tracks student interactions with course contents through page browsing activities, assessment activities (e.g., quiz), and communication activities (e.g., forum, messages). Figure 2.1 presents an architecture of our approach.

To collect the data in a central repository, HackyStat sensors are embedded within Eclipse, the MILA tutors, VPL, and Moodle. All data are sent to the HackyStat sensor base. The sensor base contains all the data collected from HackyStat sensors. The sensor base is essentially a Derby database.

To extract meaning from the data stored in the sensor base, we build a set of ontologies for each student using the incoming data. The incoming data flow continuously, and hence, the ontology is instantiated continuously as and when the data arrive. For instance, given a specific programming problem, the abstract syntax tree of the student source code will be time-stamped and stored in ontology. This ontology records the state of the abstract syntax tree of the student and the changes it undergoes continually. A second ontology captures errors made by the student during the development of the code. This ontology will store the results generated by code analysis tools such as Eclipse

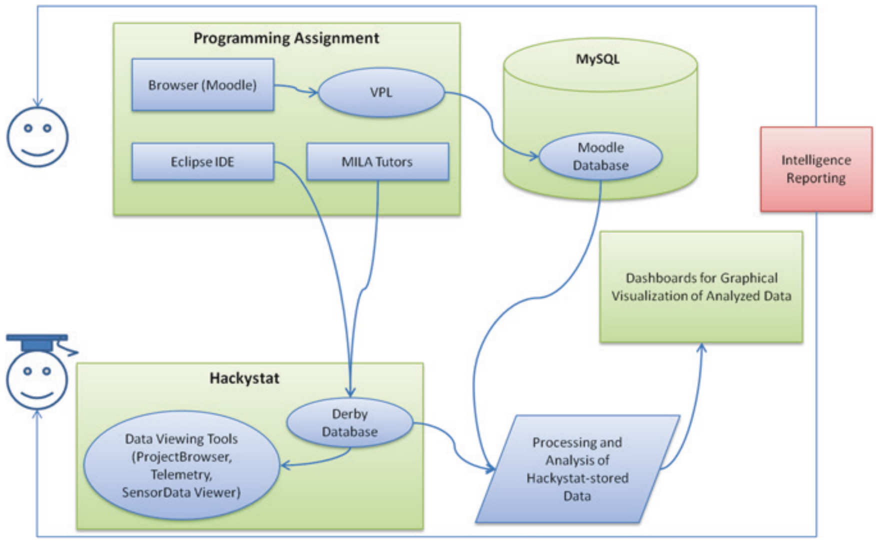

Fig. 2.1 System architecture

JDT[5]/CDT[6] compilers, FindBugs, CppCheck, and Checkstyle. To handle ontologies programmatically, we use the Apache Jena[7] framework. Ontologies will be written in the RDF or OWL format depending on the needs of expressivity or power of inference engines. We use BaseVISor,[8] an RDF-based forward-chaining inference engine, to extract useful inferences from these two ontologies. The outcome of inferences and other statistics obtained from the data stored are presented in a dashboard called MI-DASH.

2.4.1 HackyStat Sensors

HackyStat provides a range of sensors to collect data from tools such as Eclipse, Ant, Checkstyle, Clover, FindBugs, JUnit, Visual Studio, and tools generating XML data. HackyStat also provides a sensor shell Java API which supports the creation of customized sensors to increase the range of tools monitored by HackyStat sensors. The main advantage of the sensor shell is that it has a built-in data integrity mechanism; that is, the sensor sends packets of data to HackyStat at specified time intervals or when an event occurs (clicking a button, pressing the return key, etc.) via a network (be it intranet or Internet). However, if for some reason network connectivity is lost while data are being sent, the sensor shell will temporarily store the data in XML files on the user's local machine until connectivity is restored. Then, any local sensor data are automatically pushed on the HackyStat sensor base. This ensures that no data are lost due to some unexpected loss of connectivity.

Since HackyStat is open source, we modified its Eclipse sensor to suit the particular needs of our experiments. Essentially, we kept the data types originally designed with the sensor and added new ones. The sensor was also modified to fit various versions of Eclipse for various packages (Helios, Galileo, Indigo).

Finally, we created our own Eclipse package with our customized HackyStat Eclipse sensor embedded. That package also contains course assignment projects ready to be monitored by the sensor. The purpose of this package is to hide the complexity of the sensor configurations from the students.

HackyStat is designed to collect data in the background without intervening with the application. For example, once embedded in Eclipse, students would not even notice that HackyStat is collecting data. Alternatively, HackyStat can be programmed to show exactly what data are being collected, when, and how it reaches the Web server for storage. Shown in Fig. 2.2 is HackyStat's project browser that shows data being collected corresponding to a project.

[5] http://help.eclipse.org/helios/index.jsp?topic=%2Forg.eclipse.jdt.doc.isv%2Freference%2Fapi%2Foverview-summary.html.

[6] http://help.eclipse.org/indigo/index.jsp?topic=%2Forg.eclipse.cdt.doc.isv%2Freference%2Fapi%2Foverview-summary.html.

[7] http://jena.apache.org/.

[8] http://www.vistology.com/basevisor/basevisor.html.

Timestamp	Owner	SensorDataType	Runtime	Tool	Resource	Type	Subtype	Description	Username
2013-12-18 1:49:11 PM	hackystat@athabascau.ca	DevEvent	2013-12-18 1:49:11 PM	Emacs	BeijingNormalUniversity/CMP277/Assignment2/Car.java	Edit	StateChange	Generated when the active buffer has changed.	Guang Chang
2013-04-16 2:41:14 AM	hackystat@athabascau.ca	DevEvent	2013-04-16 2:41:14 AM	Emacs	ChileUniversity/INF3073/Project/Hospital.java	Edit	OpenFile	Generated when a new file is opened.	Gustavo Baloian
2013-05-09 4:06:44 PM	hackystat@athabascau.ca	DevEvent	2013-05-09 4:06:44 PM	Emacs	AnnaUniversity/MIT-C/Assignment/PetrolPump.c	Edit	SaveFile	Generated when the current file is saved.	Konijeti Rosaiah
2013-10-01 2:32:33 AM	hackystat@athabascau.ca	DevEvent	2013-10-01 2:32:33 AM	Emacs	TaiwanUniversity/Java/Assignment2/Robot.java	Edit	CloseFile	Generated when the current file is closed.	Pan-Chyr Yang
2013-10-06 5:16:32 AM	hackystat@athabascau.ca	DevEvent	2013-10-06 5:16:32 AM	Emacs	AthabascaUniversity/COMP268/TMA1/Calculator.java	Edit	!Exit	Generated when Emacs is exited	John Smith
2013-02-19 12:09:44 AM	hackystat@athabascau.ca	DevEvent	2013-02-19 12:09:44 AM	Eclipse	BeijingNormalUniversity/CMP277/Assignment2/Car.java	Edit	StateChange	Generated when the active buffer has changed.	Guang Chang
2013-02-19 11:24:04 AM	hackystat@athabascau.ca	DevEvent	2013-02-19 11:24:04 AM	Eclipse	ChileUniversity/INF3073/Project/Hospital.java	Edit	OpenFile	Generated when a new file is opened.	Gustavo Baloian
2013-07-25 10:37:19 PM	hackystat@athabascau.ca	DevEvent	2013-07-25 10:37:19 PM	Eclipse	AnnaUniversity/MIT-C/Assignment/PetrolPump.c	Edit	SaveFile	Generated when the current file is saved.	Konijeti Rosaiah
2013-03-16 6:48:51 PM	hackystat@athabascau.ca	DevEvent	2013-03-16 6:48:51 PM	Eclipse	TaiwanUniversity/Java/Assignment2/Robot.java	Edit	CloseFile	Generated when the current file is closed.	Pan-Chyr Yang
2013-06-13 4:59:43 PM	hackystat@athabascau.ca	DevEvent	2013-06-13 4:59:43 PM	Eclipse	AthabascaUniversity/COMP268/TMA1/Calculator.java	Edit	Exit	Generated when Eclipse is exited.	John Smith
2013-02-20 1:08:29 PM	hackystat@athabascau.ca	DevEvent	2013-02-20 1:08:29 PM	Eclipse	BeijingNormalUniversity/CMP277/Assignment2/Car.java	Edit	ProgramUnit(Add)	Generated when adding a new program construct such as an import statement, class, field, or method.	Guang Chang
2013-10-13 9:48:16 PM	hackystat@athabascau.ca	DevEvent	2013-10-13 9:48:16 PM	Eclipse	ChileUniversity/INF3073/Project/Hospital.java	Edit	ProgramUnit(Rename)	Generated when renaming a program construct.	Gustavo Baloian
2013-01-01 8:09:53 PM	hackystat@athabascau.ca	DevEvent	2013-01-01 8:09:53 PM	Eclipse	AnnaUniversity/MIT-C/Assignment/PetrolPump.c	Edit	ProgramUnit(Remove)	Generated when removing a program construct.	Konijeti Rosaiah

Fig. 2.2 HackyStat collects data and records it for web browsing

The Eclipse sensor enables to collect different data types during code development. The three types of data captured by the sensor are Edit, Build, and Debug. It tracks any operations made on a file such as opening, saving, or closing a file. It also records any addition, renaming, removal, and move of programming constructs. Any failed compilation is also sensed and sent to the sensor base. The sensor also reports debugging activities such as the start and termination of debugging sessions, the setting and removal of break points, and the passage of the debugger into or over a block of code. We also modified the sensor so that the source code may also be collected as the student types. The sensor collects the source code at 1-s intervals or whenever the student compiles his/her code (saves the code file) so that we might reconstruct his/her code at any time.

As for the MILA tutor, we decided to build a sensor from scratch using the sensor shell API. The MILA tutor records the student source code as he/she presses the up/down arrow keys or the return key. There are different tutors for different programming problems. Each tutor consists of a client-side Java Web Start application and a server-side Java servlet. The Java Web Start application stores data collected from the student in a buffer and sends the data over HTTP post messages to the Java servlet. The servlet contains a customized sensor which sends to the HackyStat sensor base all the instances of data types collected by the client application. The servlet also parses the final version of the student's source code to generate the corresponding abstract syntax tree and compiles the code using the

Eclipse JDT/CDT compiler. The tutors capture every source code change, build the code, and parse it to track the evolution of code and errors during problem solving. Code change data and error data are sent to the HackyStat sensor base. Source code is captured as a DevEvent sensor data type, while errors are captured as CodeIssue sensor data types.

As for the Moodle Web sensors, we note that the HackyStat sensor shell (as well as all of HackyStat) is built in Java, but since Moodle is built using PHP, JavaScript, and AJAX, we use a framework called PHP/Java Bridge which is designed to allow the integration of Java libraries inside PHP code and vice versa. Therefore, we use the sensor shell Java API inside the Moodle PHP code to build a HackyStat sensor for Moodle that uses the data integrity feature of the Java sensor shell.

We create event handlers in Moodle's code that incorporate the Web HackyStat sensors which send data about the event that just occurred (action type, time stamp, action context, any user input, etc.) to the HackyStat sensor base.

2.4.2 Data Analysis

Discussion around the analysis of the data is centered on the creation of key performance indicators of learning analytics that fall along three distinct dimensions: overall course measurement, student-based traces, and content-specific traces.

With the ability to collect formative data on student learning activities, data on challenges faced by students, data on ways in which students reacted to these challenges, data on the quality of instructional supported afforded to students, and data on the means through which effectiveness of instruction is measured could be traced and collected. These traces offer valuable information about learner capabilities, instructional effectiveness, and the overall quality of instructional environment. For example, a student profile on coding skills can be built as shown in Fig. 2.3.

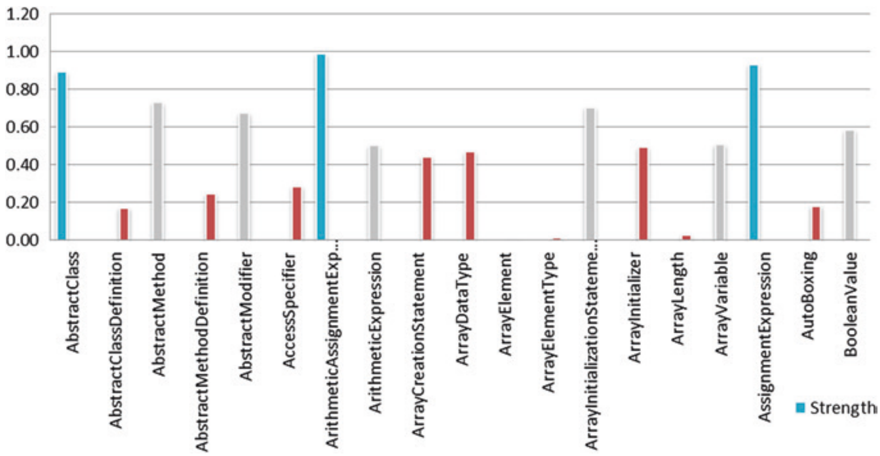

Fig. 2.3 A student profile on coding competency

What they do not offer are the means through which the instructional gaps could be addressed. For instance, if a student struggled to develop appropriate test cases, these traces provide clues on theses struggles. However, each student (or a group of students with similar issues) needs individualized instruction that adapts to his/her specific requirements. This is an opportunity for interaction with the student to address specific instructional gaps. With this in mind, we developed additional options within each tool that students could use to engage in these interaction opportunities. For instance, within each MILA tutor, we built opportunities for a type of self-regulation and co-regulation. Similarly, the Eclipse IDE itself is extended with query interfaces for students to supply data on their knowledge gaps and seek information about others who might have faced similar gaps and how they proceeded to address them. This is possible because HackyStat and its sensors continuously collate information on progress of students in specific coding activities and groups them using a particle swarm optimization algorithm.

Here, we provide some sample outcomes of the system. First, one could get simple statistics such as the number of learner interactions within Moodle with respect to the course. An interaction is the activation of any study tool within Moodle. We observed that over a period of a semester, 182 students out of the total 240 recorded 42,781 interactions within Moodle. During the same time period, these 182 students also attempted 11,151 times to write code using VPL. Eventually, these 182 students submitted 1,147 complete programs for evaluation. These 182 students also attempted to take self-study quizzes 1,693 times and viewed course content 7,631 times.

We can then dig deeper to extract study sequences of successful students. For instance, based on the marks received in the first assignment, one could gather the number of times successful students (say, those who scored 90 % of above in the assignment) attempted to solve each programming problem. We can then allow a student to compare his/her attempts with that of the successful student. That is, the student can perform a line-by-line comparison of code between each of his/her attempt to write the program with that of a successful student.

The student can also compare the time he/she took to solve a problem versus the average time of a successful student. Digging deeper, the student could investigate the exact situation that caused the difference in problem-solving time. Thus, the system provides query interfaces for participants to extract key information and then allows the participants to engage in guided, mixed-initiative reflective interactions for a deeper understanding of issues.

Further, the system allows participants to engage in regulatory activities where study strategies and instructional strategies can be recorded and followed. For example, a student participant can look at the average number of significant errors that he/she created. Based on this, he/she can investigate how long he/she took to resolve each of these errors in comparison with other standard metrics. The student can then set himself/herself a goal that allows him/her to solve errors faster. The system will continue to monitor progress made by the student and offer him/her proactive feedback on the achievement of this goal.

The sequences through which students attempt to negotiate learning can be captured using a variety of sensors. The captured data allow individual students or instructors or even the organization itself to supplement students' study activities with additional instructional tools. The learning analytics system, by itself, does not prescribe to any set of tools but allows users to include (or exclude) any tool of their choice. Each included tool will be required to collect formative data as per the specifications. Thus, our vision is not to prescribe any suite of tools but to encourage users to pick and choose their own choice of tools, as long as each tool contributes to data capture through HackyStat.

The choice of tool customization could be based on the needs. For example, with respect to content quality, one could ask—'Is there a reduction in coding errors by a student who first watches the lesson and then attempts the program versus a student who simply attempts to answer the problem (so, does watching the lesson improve the student's programming)?' Other sample questions are—'Is there any difference between a student that views the videos, attempts the question, and then does the programming exercise, versus a student who does not?' and 'Is there an improvement in coding submissions over time, as the student becomes more proficient?' As mentioned before, such questions can be posed by any participant (e.g., a student or an instructor) and answers can be received from the system in real time. A participant may choose to investigate further to explore the types of students who benefit from a particular sequence of study activities, to gain more instructional insights. Thus, the range of application of analytics goes from looking specifically into one particular competency development for one student to a trend in learning from large groups of learners.

2.5 Results and Contributions

One of the main benefits of the HackyStat/VPL/Eclispe/Moodle customizations is the ability to collect and store the raw data for later processing—and even to combine datasets over time, where appropriate, to gradually increase our total sample size over time. Thus, we are able to collect an increasing amount from a number of different classes, with almost no intrusion into the learning environment from the perspective of either the student or the instructor.

Further, as both VPL and HackyStat are multilingual and open source, we are able to deal with data from a number of different programming languages, including C, C++, Java, and Python, and even have the ability, with some additional customizations, to collect data from students studying across different languages.

The main anticipated technological contributions are twofold—one, to provide an ontology-based data mining/analytics framework, and two, a causal model for developing inferences from the relationship between learning activities and coding competency. These inferences may involve optimal learning strategies, an optimal mix of learning media, or may identify subgroups with different learning characteristics and requirements.

A secondary but equally compelling contribution is the development of a comprehensive software traces platform reporting on the effectiveness of digital introductory computer programming course material and study habits and feeding that data back to both students and instructors.

Through the combination of these two results, a third goal is the creation of an intelligent agent that can monitor the data sources on a student-by-student basis, to identify successful patterns of learning, based on a student's interaction with the digital learning environment, and can suggest corrective actions, materials, and behaviors that might improve performance, based on the past results of similar student behaviors.

Learning analytics is not about guiding students in a particular, well defined study pathway. Instead, it is about guiding students to explore meaningful and personalized study pathways that have a higher potential of success.

The data streams used in learning analytics are neither finite nor well defined. Instead, learning analytics allows any tool to offer a new data stream that can be of use. API specifications are being developed to ensure data compatibility, semantic coherence, and tool integration.

Learning analytics does not aim to offer solutions to problems encountered by students as they study. Instead, it offers students to find study pathways that might lead to solutions.

Learning analytics is mostly about the 'ING' in learnING—the process of learning and how it arrives at products and outcomes. For example, along with the review of the thesis draft submitted by a student, a supervisor could analyze the processes/pathways the student took to reach that draft. This analysis includes the articles the student referenced, passages he/she had quoted/used in the draft, passages written by the student, ideas independently generated by the student, whether references resulted in a modification of certain ideas presented in the draft, the pace of writing over the entire duration, the impact of the supervisor's discussions with the student during the development of this essay, and so on. Thus, learning analytics aims to offer as well as discover useful clues that help students aim at better learning capacity and instructors aim at effective instruction.

The technologies we have currently included range from simple statistics, to rule-based approaches to pattern recognition, to mixed-initiative conversational agents, and to causal models that discover causal connections between study pathways and competencies. The technologies are embedded in a learning analytics platform that is both extensible and open. Anyone could add a new piece of tool/technology that offers useful data to the learning analytics platform. Further, any tool/technology that has been submitted is also made available for others to use.

We are at the cusp of a significant change in perspective on learning that not only offers to guide students to specific solutions but also allows them to reflect/regulate their solution pathways in an adaptive, personalized, optimal, guidable, just-in-time fashion, and wholistic fashion—an approach that directly meets the goals of smart learning environments.

Acknowledgements We gratefully acknowledge funding support from NSERC Canada, Athabasca University, and Shastri Indo-Canadian Institute for this research.

References

Abdous, M., He, W., & Yen, C. J. (2013). Using data mining for predicting relationships between online question theme and final grade. *Educational Technology and Society, 15*(3), 77–88.

Allen, J. F., Guinn, C. I., & Horvitz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems, 14*(5), 14–23.

Angelo, T. A., Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers,* (2nd ed.). San-Francisco: Jossey-Bass.

Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting collaborative learning and problem-solving in a constraint-based CSCL environment for UML class diagrams. *International Journal of CSCL, 2*(2–3), 150–190.

Bienkowski, M., Brecht, J., & Klo, J. (2012). The learning registry: Building a foundation for learning resource analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 208–211).

Boekaerts, M., Corno L. (2005). Self-regulation in the classroom: A perspective on assessment and interventions. *Applied Psychology: An International Review, 54*, 199–231.

Brooks, C., Winter, M., Greer, J., & McCalla, G. (2004). The massive user modelling system. In *The Proceedings of the 7th International Conference on Intelligent Tutoring Systems, Maceio, Brazil* (pp. 635–645).

Butler, D. L., Schnellert, L., & Higginson, S. (2008). *Fostering agency and co-regulation: Teachers using formative assessment to calibrate practice in an age of accountability*. Paper presented at the American Educational Research Association, April 2008.

Cross, K. P., Angelo, T. A. (1988). *Classroom assessment techniques: A handbook for faculty*. Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Dicheva, D., Mizoguchi R., & Greer J. (2009). *Semantic web technologies for e-learning*. IOS Press, ISBN 978-1-60750-062-9.

Essa, A., & Ayad, H. (2012). Improving student success using predictive models and data visualisations. In *Proceedings of the ALT-C 2012 Conference* (pp. 58–70).

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning (IJTEL), 4*(5/6), 304–317.

Ghidini, C., Pammer, V., Scheir, P., Serafini, L., & Lindstaedt, S. (2007). APOSDLE: Learn@work with semantic web technology. In *Proceedings of the International Conference on Knowledge Management (I-Know '07), Graz, Austria*.

Hung, J. L., Hsu, Y. C., & Rice, K. (2012). Integrating data mining in program evaluation of K-12 online education. *Educational Technology and Society, 15*(3), 27–41.

Jovanović, J., Gašević, D., Torniai, C., Bateman, S., & Hatala, M. (2009). The social semantic web in intelligent learning environments-state of the art and future challenges. *Interactive Learning Environments, 17*(4), 273–308.

Kim, M., & Lee, E. (2013). A multidimensional analysis tool for visualizing online interactions. *Educational Technology and Society, 15*(3), 89–102.

Kosba, E., Dimitrova, V., Boyle, R. (2005). Using student and group models to support teachers in web-based distance education. In *Proceedings of the 10th International Conference on User Modeling, Edinburgh, UK* (pp. 124–133).

Manyika, J., Chui M., Brown, B., Bughin, J., Dobbs, R., & Roxburgh, C. (2011). *Big data: the next frontier for innovation, competition, and productivity*. Report by McKinsey Global Institute. Retrieved January 10, 2014, from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Mazza, R., Dimitrova, V. (2003). CourseVis: Externalising student information to facilitate instructors in distance learning. In *Proceedings of the International Conference in Artificial Intelligence in Education (AIED 2003)* (pp. 279–286).

Mazza, R., Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. In *Proceedings of the 13th World Wide Web Conference, NY, USA* (pp. 154–161).

Mazza, R., Milani, C. (2005). Exploring usage analysis in learning systems: gaining insights from visualisations. In *Proceedings of the Workshop on Usage analysis in learning systems at the 12th International Conference on Artificial Intelligence in Education, Amsterdam, The Netherlands* (pp. 65–72).

Medeiros, F., Gomes, A. S., Amorim, R., & Medeiros, G. (2013). Architecture for social interactions monitoring in collaborative learning environments as a support for the teacher's awareness. In *Proceedings of International Conference on Advanced Learning Technologies (ICALT)* (pp. 123–127).

Rao, S., Kumar, V., Hatala, M., Gašević, D. (2007). Mixed-initiative interfaces to recognize regulate and reflect programming styles. In *Proceedings. of the Workshop on AI for Human Computing at 20th International Joint Conference on Artificial Intelligence (IJCAI), India* (pp. 71–78).

Romero, C., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle cases study and tutorial. *Computers and Education, 51*(1), 368–384.

Rountree, N., Rountree, J., Robins, A., Hannah, R. (2004). Interacting factors that predict success and failure in a CS1 course. *ACM SIGCSE Bulletin, 36*(4), 101–104.

Shakya, J. (2005). *Knowledge engineering and knowledge dissemination in a mixed-initiative ontological framework*. MSc thesis, Simon Fraser University, Canada.

Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Educational Technology and Society, 15*(3), 3–26.

Watson, C., Li, F. W. B., & Godwin, J. L. (2013). Predicting performance in an introductory programming course by logging and analyzing student programming behavior. In *Proceedings of International Conference on Advanced Learning Technologies (ICALT)* (pp. 319–323).

Wood, E., Zivcakova, L., Gentile, P., Archer, K., Pasquale, D. D., & Nosko, A. (2013). Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers and Education, 58*(2011), 365–374.

Zinn, C., and Scheuer, O. (2006). Getting to know your student in distance-learning contexts. In *Proceedings of the 1st European Conference on Technology Enhanced Learning* (pp. 437–451).

Chapter 3
Semantic Similarity-Enhanced Topic Models for Document Analysis

Yan Gao and Dunwei Wen

Abstract In e-learning environment, more and more larger-scale text resources are generated by teaching–learning interactions. Finding latent topics in these resources can help us understand the teaching contents and the learners' interests and focuses. Latent Dirichlet allocation (LDA) has been widely used in many areas to extract the latent topics in a text corpus. However, the extracted topics cannot be understood by the end user. Adding more auxiliary information to LDA to guide the process of topic extraction is a good way to improve the interpretability of topic modeling. Co-occurrence information in corpus is such information, but it is not sufficient yet to measure the similarity between word pairs, especially in sparse document space. To deal with this problem, we propose a new semantic similarity-enhanced topic model in this paper. In this model, we use not only co-occurrence information but also the semantic similarity based on WordNet as auxiliary information. Those two kinds of information are combined into a topic-word component though generative Pólya urn model. The distribution of documents over the extracted topics obtained by the new model can be inputted to the classifier. The accuracy of extracting topics can improve the performance of the classifier. Our experiments on newsgroup corpus show that the semantic similarity-enhanced topic model performs better than the topic models with only single information separately.

Keywords Topic modeling · LDA · Gibbs sampling · Generative Pólya urn model · Semantic similarity · WordNet

---

Y. Gao
School of Information Science and Engineering, Central South University, Changsha, Hunan, China

Y. Gao · D. Wen (✉)
School of Computing and Information Systems, Athabasca University, Athabasca, AB, Canada
e-mail: dunweiw@athabascau.ca

Y. Gao
Hunan Engineering Laboratory for Advanced Control and Intelligent Automation, Changsha, Hunan, China

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_3

3.1 Introduction

With the development of information technology, more and more large-scale text repositories have been established. Mining a large-scale text repository to find latent knowledge and information is a difficult task. Latent Dirichlet allocation (LDA) is a statistical topic model proposed by Blei et al. (2003). It can automatically extract semantic themes from large, unstructured collections of text documents. It has been applied in a wide range of areas: analysis of news articles (Newman et al. 2006), social analysis of microblog, such as on twitter (Ramage et al. 2010), study of the history of scientific ideas (Griffiths and Steyvers 2004; Rosen-Zvi et al. 2004), topic-based ad hoc document retrieval (Zhai and Lafferty 2001), text sentiment analysis (Lin and He 2009), clustering of digital libraries content (Mimno and McCallum 2007; Wang et al. 2012), word sense disambiguation (Boyd-Graber et al. 2007), and information extraction (Zhang et al. 2009).

However, the topics extracted by LDA are not interpretable or meaningful. Some of the topics can be a collection of irrelevant or background words, and some represent insignificant themes. One may feel hard to understand or figure out the very topics that the extracted words represent. This problem is more serious in the content that is sparse or noisy, such as blog posts, tweets, or e-mail than traditional documents. There are two reasons to illustrate these phenomena. One is that the objective function of topic models does not always correlate with human judgments (Chang et al. 2009). The other is that topic models are based on a "bag of words" approach, which lack the necessary information to create the topics as end users expect.

In order to overcome this shortcoming, researchers focus on adding more auxiliary information to topic models: meta information such as authorship (Rosen-Zvi et al. 2004), domain knowledge on words (Must-Link and Cannot-Links) (Andrzejewski et al. 2009, 2011; Chen et al. 2013), corpus' co-document frequency information (Mimno et al. 2011; Newman et al. 2011), and user's guiding information that enforces that sets of words must appear together in the same topic (Chen et al. 2013).

In e-learning environment, students communicate with teachers by e-mail, forums, and microblogs etc. By a quick analysis of these sources of texts, we can find that they are informal, sparse, and time-varied. The information from co-document frequency is not sufficient. To handle these kinds of texts, this chapter will present a new topic model: semantic similarity-enhanced topic model (SETM) in which the lexical semantic similarity based on WordNet is used to guide topic modeling besides co-document frequency information. We first provide some key techniques/components of topic modeling in Sect. 3.2 and then introduce the details of the semantic similarity-enhanced topic model in Sect. 3.3, and we finally describe the experiments in Sect. 3.4 to demonstrate the new topic model's effectiveness.

3.2 Topic Modeling

Topic modeling is a type of unsupervised machine learning technique which generates latent topics from text documents. Latent Dirichlet allocation (LDA) is the most common topic model currently in use. Comparing with other topic modeling methods such as LSI and pLSI, LDA has good performance, which can probabilistically assign a group of similar words into the same topic. In this section, we briefly introduce the theory of LDA and the inference algorithm for LDA.

3.2.1 Latent Dirichlet Allocation (LDA)

Latent Dirichlet allocation (LDA) is a three-level hierarchical Bayesian model. In LDA, each document in a corpus is modeled as a multinomial distribution over an underlying set of K topics, and each topic is characterized by a multinomial distribution over V words (Blei et al. 2003). Figure 3.1 shows the graphical representation of LDA in plate notation.

Given the words of M documents containing K topics is generated from V unique words, the generative process for each document d in a corpus $D = (d_1, d_2, \ldots d_M)$ defined by LDA is shown in Table 3.1.

3.2.2 Gibbs Sampling for LDA

Although LDA is still a relatively simple model, exact inference is generally intractable. Blei et al. (2003) defined the variational inference algorithm for training topics. Griffiths proposed the collapsed Gibbs sampling algorithm (Canini et al. 2009; Griffiths and Steyvers 2004) to perform approximate inference for LDA model. Minka proposed exception propogation (Minka 2001) for LDA. Comparing with the variational inference algorithm and approximate inference, the Gibbs sampling algorithm for LDA is favored for its simplicity and speed.

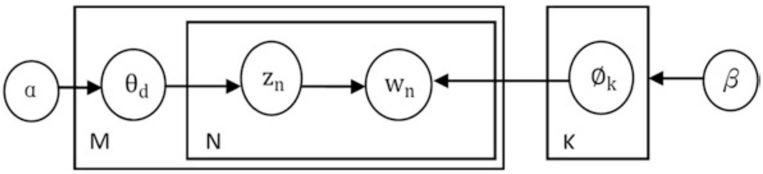

Fig. 3.1 Graphical representation of LDA in plate notation (Blei et al. 2003)

Table 3.1 The generative process for each document in LDA

The generative process for each document in LDA
1. For each topic $k \in (1,\ldots K)$,
Draw a multinomial distribution ϕ_k from a Dirichlet prior β, where ϕ_k is the word distribution of topic k;
2. For each document d_m, $m \in (1,\ldots, M)$:
(a) Draw a multinomial distribution ϑ_m from a Dirichlet prior α, where ϑ_m is the topic distribution for document m (b) For each word $n \in (1,\ldots, N_m)$ in document m:
i. Draw a topic $z_{m,n}$ from multinomial ϑ_m, where $z_{m,n}$ is the topic for the nth word in document m
ii. Draw a word $w_{m,n}$ from multinomial $\phi_{zm,n}$, where $w_{m,n}$ is the specific word in document m

The whole process of collapsed Gibbs sampling algorithm (CGS) is to iteratively sample a topic z for every word in every document from a full conditional probability. The sampling update formula is defined as:

$$P\left(z_{m,n}|\vec{w},\vec{z_{\neg(m,n)}},\alpha,\beta\right) \propto \frac{n_{w_{m,n}|z_{m,n}}+\beta-1}{\sum_{v=1}^{V}\left(n_{v|z_{m,n}}+\beta\right)-1}\times\left(n_{z_{m,n}|m}+\alpha\right) \tag{3.1}$$

where $z_{m,n}$ is the topic assigned to nth word in document m:$w_{m,n}$.

$n_{v|z_{m,n}}$ is the number of tokens of word v are assigned to topic $z_{m,n}$,
$n_{z_{m,n}|m}$ represent the number of tokens in document m are assigned to topic $z_{m,n}$.
$\overline{z_{\neg(m,n)}}$ represents the topic assignments for all tokens except $w_{m,n}$.

After sampling process is finished, the topics ϕ_k and the mixture of topics for documents θ_m can be learned simultaneously as follows (Heinrich 2009),

$$\phi_{k,v}=\frac{n_{v|k}+\beta}{\sum_v(n_{v|k}+\beta)} \tag{3.2}$$

$$\theta_{m,k}=\frac{n_{k|m}+\alpha}{\sum_k\left(n_{k|m}+\alpha\right)} \tag{3.3}$$

where $n_{k|m}$ is the number of tokens of word v that are assigned to topic k, $n_{\ k|m}$ represents the number of tokens in document m that are assigned to topic k. $\phi_{k,v}$ is the probability of topic k belonging to word v. $\theta_{m,k}$ is the probability of document m belonging to topic k.

Using Eqs. (3.1), (3.2), and (3.3), the Gibbs sampling procedure in Table 3.2 (Mimno et al. 2011) can be run iteratively until convergence or the L iteration time is reached. The main steps in Gibbs sampling include initialization and sampling.

Because LDA is based on "bag of words" approach, the "bag of words" representation of document is the input of CGS algorithm. The other parameter inputs include α, β, and the topic number K.

Table 3.2 Gibbs sampling algorithm for LDA

```
Gibbs sample Algorithm for LDA
//Initialization
   Zero all count variable: n_{k|m}, n_{v|k}
   For all document d_m
        for all words w_{m,n} ∈ d_m
             random sampling topic index z_{m,n} = k ~ Mutinomial(K)
             increase document-topic count: n_{k|m} +=1
             increase topic-word count: n_{v|k} +=1

// Gibbs sampling
  while not finished do

     // one sweep of LDA Gibbs sampling (Mimno et al. 2011)
     for d_m ∈ D, M=[d_1,......d_M]
          for w_{m,n} ∈ d_m do

             Z_i = Z_{m,n}

             N_{z_i|d} ← N_{z_i|d} − 1

             N_{w_{m,n}|z_i} ← N_{w_{m,n}|z_i} − 1

             Sampling z_i ∝ (N_{z_i|d} + α_z) (N_{w_{m,n}|z_i} + β) / (Σ_{v=1}^{V} (n_{v|z_i} + β))

             N_{z_i|d} ← N_{z_i|d} + 1

             N_{w_{m,n}|z_i} ← N_{w_{m,n}|z_i} + 1

         end for
     end for
     //one sweep of LDA Gibbs sampling ends

     if converged or L iterations
        for k ∈ [1..K]
          compute ϕ_k using Eq (3.2)
          for d_m ∈ D, m ∈ [d_1,......d_M]
               compute ϑ_m using Eq (3.3)
```

3.3 Semantic Similarity-Enhanced Topic Model

In this section, we propose a semantic similarity-enhanced topic model. In this model, we employ the generalized Pólya urn (GPU) model for adding semantic similarity to LDA. The generalized Pólya urn (GPU) model is introduced in

Sect. 3.1 in brief. And our semantic similarity measurement based on co-occurrence information and lexical semantic information based on WordNet are introduced in Sect. 3.2.

3.3.1 Generative Pólya Urn Model

Pólya urn model is a renowned statistical model. In Pólya urn model, colored balls in an urn are used to represent objects of real interest (such as words and topics). A ball is sampled at random from the urn. After the color of drawing ball is observed, it is put back, together with another ball of the same color (Mahmoud 2008). This sampling process is iteratively repeated. Pólya urn model is thus a statistical model of sampling with replacement. When sampling with replacement, the observed color is much likely to be observed again. The topic-word component of LDA is a Pólya urn model. Given W and Z, the conditional posterior probability of word w in topic z generated by LDA is as follows:

$$P(w|t, W, Z, \beta, A) = \frac{N_{w|t} + \beta}{\sum_v N_{v|t} + V\beta} \tag{3.4}$$

Generative Pólya urn model (GPU) (Mahmoud 2008) changes the replacement policy in the sampling process of Pólya urn model. In generative Pólya urn model, A_{vw} additional balls of each color are put back in the urn after randomly sampling a ball. Mimno et al. (2011) introduced the generative Pólya urn model into LDA model. In his new model, Pólya urn model, which controls the topic-words component, is replaced with generalized Pólya urn model. So the occurrence of word type w in topic t increases the probability of seeing that word type again and the probability of seeing other related words simultaneously.

Given W and Z, the conditional posterior probability of word w in topic z generated by GPU is as follows (Mimno et al. 2011):

$$P(w|t, W, Z, \beta, A) = \frac{\sum_v N_{v|t} A_{vw} + \beta}{\sum_v N_{v|t} + V\beta} \tag{3.5}$$

where A_{vw} is the relatedness value between word w and v. A is row-normalized to 1.

According to Eqs. (3.4) and (3.5), we can modify the Gibbs sampling algorithm listed in Table 3.2 to obtain the new Gibbs sampling algorithm for generative Pólya urn mode. Because A is normalized, we still use Eqs. (3.2) and (3.3) to compute ϕ_k and θ_m. Focusing on the difference between the two algorithms, we only describe a single Gibbs sampling for LDA with generative Pólya urn model in Table 3.3 (Mimno et al. 2011).

Table 3.3 A single Gibbs sampling for LDA with generative Pólya urn model (Mimno et al. 2011)

Algorithm for a single Gibbs sampling for LDA with generative Pólya urn model

For $d_m \in D$, $m \in (d_1, \ldots \ldots d_M)$
    for $w_{m,n} \in d_m$ do
        $Z_i = Z_{m,n}$
        $N_{z_i\|d} \leftarrow N_{z_i\|d} - 1$
        For all v do
            $N_{v\|z_i} \leftarrow N_{v\|z_i} - A_{vw}$
        Sampling $Z_i \propto (N_{z_i\|d} + \alpha_z) \frac{N_{w_{m,n}\|z_i} + \beta}{\sum_{v=1}^{V} (n_{v\|z_i} + \beta)}$
        $N_{z_i\|d} \leftarrow N_{z_i\|d} + 1$
        For all v do
            $N_{v\|z_i} \leftarrow N_{v\|z_i} + A_{vw}$
    end for
end for

3.3.2 Semantic Similarity-Enhanced Topic Model

As above mentioned, A_{vw} is the important parameters in generative Pólya urn model. In this section, we propose a new measurement which combines semantic similarity based on WordNet and co-occurrence information to define A_{vw}.

Commonly, the relatedness between word pair v and w is decided by the degree of semantic similarity. Co-document frequency between word pair in a corpus is a way to measure semantic similarity between two words. Mimno used the corpus-specific word co-occurrence information to measure A_{vw} (Mimno et al. 2011). A_{vw} is defined as:

$$A_{vv} \propto \lambda_v D(v) \tag{3.6}$$

$$A_{vw} \propto \lambda_v D(w, v) \tag{3.7}$$

$$\lambda_v = \log(D/D(v)) \tag{3.8}$$

where $D(v)$ is the document frequency of word v, and $D(w,v)$ is co-document frequency of word w and v. The standard IDF weight λ_v is used to weight row-specific elements, which can decrease the effect of the more frequent word v.

With the development of web 2.0, more and more student and teachers use forum, twitter, e-mail, and question answering system to establish their communication. Mining these texts can tell us many aspects of e-learning, such as teaching content and student understanding. As mentioned above, these texts have same characteristics: short, sparse, and informal.

The co-document information in short text contains less sufficient information than that in long text. The word pairs that co-occur in the short text corpus are less than those in long text corpus. Some related words do not occur in a document. So we propose a new semantic similarity-enhanced topic model (SETM) which modifies Mimmo's model by adding the semantic similarity obtained from WordNet to deal with the short and sparse document.

WordNet is a famous lexical database for English (Miller et al. 1990). The words are grouped into synsets and organized into a hypernym/hyponym hierarchy (nouns) or hypernym/troponym hierarchy (verbs). Based on WordNet, two kinds of measurements were developed to evaluate the semantic similarity between synset pairs (i.e., sense-specified word pairs). One is based on path lengths, which includes lch (Leacock and Chodorow 1998), wup (Wu and Palmer 1994), and path. The other is based on information content, which includes res (Resnik 1995), lin (Lin 1998), and jcn (Jiang and Conrath 1997). In this paper, we use path distance to evaluate the semantic similarity between synset pairs. The path distance between two nodes is defined as the number of nodes that lie on the shortest path between two words in the hierarchy. The shorter the path between two words/senses in a thesaurus hierarchy graph, the more similar they are. Path-based similarity $\mathrm{sim}^{(\mathrm{path})}$ often involves a log transform.

$$\mathrm{sim}^{(\mathrm{path})}(c_1, c_2) = -\log \mathrm{pathlen}(c_1, c_2) \tag{3.9}$$

$$\mathrm{pathlen}(c_1, c_2) = \text{number of nodes in shortest path} \tag{3.10}$$

Path-based similarity $\mathrm{sim}^{(\mathrm{path})}$ is only used to measure two synsets c_1 and c_2. The semantic similarity of two words v and w is the maximum path-based similarity among all the sense belonging to w and v:

$$A_{vw}^{(\mathrm{path})} = \max\nolimits_{c_1 \in \mathrm{synsets}(w), c_2 \in \mathrm{synset}(v)} \mathrm{sim}^{(\mathrm{path})}(c_1, c_2) \tag{3.11}$$

where synsets(w) is the collection of all senses belonging to w which must be noun or verb.

Using WordNet, we only compute the lexical semantic similarity between nouns or verbs. Some other similarity between two different parts-of-speech (POS), such as noun and adjective, or verb and adverb, cannot be obtained from WordNet. Besides the problem mentioned above, WordNet cannot contain new words and informal words which often appear in the documents on Web.

To overcome this problem, we establish a new measurement which combines co-occurrence information A_{vw} and the path distance similarity based on WordNet $A_{vw}^{(\mathrm{path})}$ to evaluate the semantic similarity of two words w and v. In our generative Pólya urn mode, A_{vw} is replaced with $A_{vw}^{(\mathrm{new})}$. $A_{vw}^{(\mathrm{new})}$ is defined as:

$$A_{vw}^{(\mathrm{new})} = \lambda \times A_{vw} + (1 - \lambda) \times A_{vw}^{(\mathrm{path})} \tag{3.12}$$

Before computing $A_{vw}^{(\mathrm{new})}$, A_{vw} and $A_{vw}^{(\mathrm{path})}$ should be column-normalized to 1. When $\lambda = 1$, the semantic enhanced topic is equal to Mimmo's model.

3.4 Experiments

In this section, we compared the experimental evaluation of semantic similarity-enhanced topic model (SETM), Mimmo's Model on the mini-Newsgroup which can be obtained from UCI Web site: http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.html. The information about the corpus is listed in Table 3.4.

Preprocessing: We ran the NLTK (Bird et al. 2009) to perform lemmatization and POS tagging. Then, punctuations, stop words, numbers, and words appearing less than 3 times in corpus were removed. The similarity based on WordNet was calculated by the software ws4j which can be downloaded from: http://code.google.com/p/ws4j/. In order to decrease the effect of words with high document frequency, we removed the off-diagonal elements of A_{vw} for words whose value λ_v was less than 2.4 (only 5 % documents in dataset).

Parameters Setting: We empirically set $\lambda = 0.5$, topic number $K = 100$, $\alpha = 1/K$, and $\beta = 1/K$ for the purpose of demonstrating the mechanisms of our framework. We ran collapsed Gibbs sampling for 500 iterations.

Topic models are often evaluated using perplexity on held-out test data. However, the perplexity measure does not reflect the semantic coherence of individual topics learned by a topic model (Newman et al. 2010). Recent research has suggested that human judgments can sometimes be contrary to the perplexity measure (Chang et al. 2009). But the human judgments are hard to be obtained. So we turned to use classifier to evaluate the new model automatically. If the latent topics can be extracted correctly by LDA, the appropriate distribution of documents over the topics can be achieved. Here, we used the task of classification to evaluate the distribution of documents over the extracted topics. The classifier used in this paper is SVM.

We use 5-fold cross-validation to evaluate the accuracy of every model. In 5-fold cross-validation, the data are first partitioned into 5 equally (or nearly equally) sized subsets. Subsequently, 5 iterations of training and validation are performed. In each iteration, a different subset is held out for testing while the remaining 4 subsets are used for training. The model SETM with $\lambda = 0$ is called path-based LDA in our experiments. The results of every iteration on our model (SETM), Mimmo's Model ($\lambda = 1$), and path-based LDA on mini-Newsgroup are listed in Table 3.5.

Table 3.4 The details of mini-Newsgroup

Number of documents	Number of words before preprocessing	Number of words after preprocessing	The proportion of co-document word pair
2,000	30,961	7,734	4.93 %

Table 3.5 The results of every iteration on our model (SETM), Mimmo's Model and path-based LDA on mini-Newsgroup

	1	2	3	4	5	Average accuracy (%)
Minnio's model + SVM ($\lambda = 1$)	0.4025	0.3275	0.3625	0.375	0.3925	37.2
SETM + SVM ($\lambda = 0.5$) (%)	49.5	48	47	48	42.5	47
Path-based LDA ($\lambda = 0$)	0.375	0.34	0.3575	0.3375	0.375	35.7

From Table 3.5, we can see that the co-occurrence information or WordNet-based semantic information used in LDA separately degrade the accuracy of topic model. The accuracy of SETM is better than Minnio's model and path-based LDA. The dataset is composed of 2,000 documents, which are all collected from newsgroups on the Web. They have the same characteristics as other text on the Web (such as e-mail and forum): sparse, noisy, and informal. Before preprocessing, only 4.93 % word pairs in the corpus have co-document relationship. The co-occurrence information is not sufficient, and the accuracy of this model is not the best but better than path-based LDA. WordNet-based semantic information only catches the semantic similarity between nouns and verbs existing in the hierarchy, ignoring other sense words and informal words not included in the dictionary. Path-based LDA is the worst among three model. In SETM, we used the combination of the co-occurrence information and the WordNet-based semantic information in LDA, so the performance is the best among three models.

3.5 Conclusion

In e-learning environment, the ease of publishing and accessing information makes the growth of learning repositories rapidly. And the more and more documents are generated by the convenient communication between students and teachers. The task of extracting latent topics can help people understand the knowledge in documents and organize the document quickly.

This chapter has presented an innovative approach for performing text mining on documents, which serves as a basis for knowledge extraction in e-learning environments. In this approach, the latent topics can be extracted by a novel semantic similarity-enhanced topic model that introduces lexical semantic similarity into topic modeling in order to produce more coherent topics. The new semantic similarity combines semantic similarity from both WordNet and co-occurrence information. Through generative Pólya urn model, the new semantic similarity is incorporated into LDA models. The distribution of documents over latent topics is used in the SVM classifier to improve the ability to classify the documents. In future, we will further investigate the effect of other similarity measurement on LDA.

Acknowledgments The authors acknowledge the support of Research Incentive Grant (RIG) of Athabasca University. This work was also partly supported by the National Natural Science Foundation of China under Grant 61273314 and 61175064.

References

Andrzejewski, D, Zhu, X. J., & Craven, M. (2009). Incorporating domain knowledge into topic via Dirichlet forest priors. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*.

Andrzejewski, D, Zhu, X. J., Craven, M., & Recht, B. (2011). A framework for incorporating general domain knowledge into latent Dirichlet allocation using first-order logic. In *Proceedings of the 22nd International Joint Conferences on Artificial Intelligence (IJCAI 2011)*.

Banerjee, S., & Pedersen, T. (2003). Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence* (pp. 805–810).

Bird, S., Loper, E., & Klein, E. (2009). *Natural language processing with python*. Sebastopol: O'Reilly Media Inc.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *JMLR, 3*, 993–1022.

Boyd-Graber J, Blei D, & Zhu X. (2007). A topic model for word sense disambiguation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (pp. 1024–1033).

Canini, K. R., Shi, L., & Griffiths T. L. (2009). Online inference of topics with latent Dirichlet allocation. In *AISTATS* (vol. 5).

Chang, J., Boyd-Graber, J., Gerrish, S., Wang C., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems (NIPS)* (vol. 21).

Chen, Z. Y., Mukherjee, A, Liu., B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Discovering coherent topics using general knowledge. In *Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management* (pp. 209–218).

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl. 1), 5228–5235.

Heinrich, G. (2009). Parameter estimation for text analysis, Technical report, Fraunhofer IGD, Available at http://www.arbylon.net/publications/text-est2.pdf

Jiang, J., & Conrath, D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings on International Conference on Research in Computational Linguistics* (pp. 19–33).

Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. In C. Fellbaum, (Ed.), *WordNet: An electronic lexical database* (pp. 265–283). Cambridge: MIT Press.

Lin, D. (1998). An information-theoretic definition of similarity. In *Proceedings of the International Conference on Machine Learning*.

Lin, C. & He, Y. (2009). Joint sentiment/topic model for sentiment analysis. In *The 18th ACM Conference on Information and Knowledge Management (CIKM)*.

Mahmoud, H. (2008). *Pólya urn models*. Chapman & Hall/CRC Texts in statistical science.

Miller, G. A., Beckwith, R., Fellbaum, C. D., Gross, D., & Miller, K. (1990). WordNet: An online lexical database. *International Journal of Lexicography, 3*(4), 235–244.

Mimno, D. & McCallum, A. (2007). Organizing the OCA: Learning faceted subjects from a library of digital books. In *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries* (pp. 376–385).

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *EMNLP* (pp. 262–272).

Minka T. P. (2001). Expectation propagation for approximate Bayesian inference. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI2001)*.

Newman, D., Chemudugunta, C., Smyth, P., & Steyvers, M. (2006). Analyzing entities and topics in news articles using statistical topic models. In *Intelligence and security informatics. Lecture Notes in Computer Science*.

Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010). Evaluating Topic Models for Digital Libraries, In: *Proceedings of the 10th Annual Joint Conference on Digital Libraries* (pp. 215–224).

Newman, D., Bonilla, E. V. & Buntine, W. L. (2011). Improving topic coherence with regularized topic models. *NIPS* (pp. 496–504).

Ramage, D, Dumais, S. T., & Liebling, D. J. (2010). Characterizing microblogs with topic models. *ICWSM 2010*.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 448–453).

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (pp. 487–494).

Wang, L. D., Wei, B. G., & Yuan, J. (2012). Document clustering based on probabilistic topic model. *Acta Electronic Sinica, 40*(11), 2346–2350.

Wu, Z. & Palmer, M. (1994). Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics* (pp. 133–138).

Zhai, C., & Lafferty, J. (2001). A study of smoothing methods for language models applied to ad hoc information retrieval. In *Proceedings of ACM SIGIR* (pp. 334–342).

Zhang, H. B., Zhu, M. J., Shi, S. M., & Wen, J. R. (2009). Employing topic models for pattern-based semantic class discovery. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL'09)*.

Chapter 4
Social Context Analysis for Topic-Specific Expert Finding in Online Learning Communities

Yanyan Li, Shaoqian Ma and Ronghuai Huang

Abstract With increasing trend toward lifelong learning, online learning communities have become important places for people to seek and share knowledge. Yet with the increasing number of members and produced artifacts within the learning communities, it is challenging to find the related documents and influential experts who post topic-specific high-quality content. In the context of Web2.0, Web documents, e.g., blogs or forum messages, are freely discussed and commented by users. The users-generated content (e.g., comments) and activities (e.g., forward) implicitly reflect the importance of Web documents. Therefore, the social context, integrating the document content information and the social event information, are analyzed in this chapter to discover expertise in online learning communities. In addition to computing documents' topic-focus degree, the proposed approach measures the quality of documents according to users' feedback behaviors, review sentiment, and topic-specific influence of users who give feedback. Experiments on real dataset have shown that our approach is effective to find the meaningful topic-specific expertise. In online learning communities, this approacth could be used to identify and recommend experts with high expertise and influence to make community interconnected and cohesive.

Y. Li (✉) · S. Ma · R. Huang
R&D Center for Knowledge Engineering, School of Educational Technology,
Beijing Normal University, Beijing, China
e-mail: liyy@bnu.edu.cn

S. Ma
e-mail: litchimsq@gmail.com

R. Huang
e-mail: huangrh@bnu.edu.cn

Y. Li · R. Huang
Collaborative and Innovative Center for Educational Technology (CICET),
Beijing Normal University, Beijing, China

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_4

Keywords Expertise finding · Online learning community · Sentiment analysis · Social context · Topic-specific

4.1 Introduction

With the widespread adoption of social media, online learning communities are perceived as a network of knowledge comprised of interconnected individuals and enormous amounts of artifacts. Though people are used to depending on search engines to search and locate useful information to solve their problems, sometimes it can be difficult for people to get satisfactory answer by searching Google directly. Instead, he or she may prefer to find and ask someone who has related expertise or experience, and online learning communities have emerged as one of the most important places for people to seek advice or help. So, the problem to find a person with a high degree of a skill or knowledge of a certain subject is an interesting and challenging issue discussed recently (Lappas et al. 2011).

Expert finding, or called expert identification, aims at answering this question "who is expert on topic X?" by analyzing interaction among people. The popularity of Web 2.0 allows people to write blogs about their activities, share knowledge in forums, write Wiki pages, and utilize social services to stay in touch with other people (Schall 2012). It is possible to utilize the knowledge of a massive amount of people participating in interactions on the Web.

Previous studies on expert finding can be classified into graph connectivity-based approaches and semantics-based topic modeling approaches (Daud 2010). Assuming that interactive connections among people can be useful to infer expertise, the graph connectivity-based approaches analyze social links by using keyword matching. On the other hand, topic modeling approach makes use of latent topic layer to semantically capture relationship. In fact, both of graph links and textual information associated with people are important for expertise finding in online learning communities. Most of the existing approaches consider only one aspect but ignore the other. With the popularity and diversity of online learning communities, social context is becoming more and more complicated. There are users, user-generated contents, social networks, and implicit networks (such as forward/reply network). So, expert finding in online learning communities depends on the social context information (e.g., social influence between users, user-generated content). Additionally, the approaches of expertise finding need to adapt to the communities' underlying interaction dynamics.

Therefore, this chapter presents a novel approach to identify expertise in online communities based on social context information. Expertise rank is measured through combination of documents' topic-focus degree, quality of documents, and users' activities. Experiment was conducted to evaluate effectiveness of the proposed approach using real dataset from Sina blog in China.

The remainder of this chapter is organized as follows. Section 4.2 discusses the related work; Sect. 4.3 presents the proposed expert finding approach; Sect. 4.4 introduces the details about the experiments and evaluation; Sect. 4.5 discusses the findings and implications in educational scenario; and Sect. 4.6 concludes the chapter.

4.2 Related Works

4.2.1 Expert Finding

Expertise finding has long been of great importance in information retrieval and filtering and has garnered significant attention in online communities. A wide range of works have been done for expertise finding. There are some expertise finder systems that help to find people with the appropriate expertise to answer a question proposed by the asker (Zhang et al. 2007). These systems have been explored in a series of studies, including Streeter and Lochbaum (1988), Krulwich and Burkey (1996), and McDonald and Ackerman (1996) as well as the studies in Ackerman et al. (2003). Newer systems, such as Yenta (Foner 1997) and ReferralWeb (Kautz et al. 1997), have also been explored by using a social network to help find people. These systems aim to leverage the social network within an organization or community to help find the appropriate others.

Most current systems use modern information retrieval techniques to discover expertise from implicit or secondary electronic resources. A person's expertise is usually described as a term vector and is used later for matching expertise queries using standard IR techniques. The result usually is a list of related people with no intrinsic ranking order or ranks derived from term frequencies. It may reflect whether a person knows about a topic, but it is difficult to distinguish that person's relative expertise levels. Relying on word and document frequencies has proven to be limited (Littlepage and Mueller 1997).

In the past, the major frameworks used for expert finding can be divided into two main categories: topology-based approaches (Erten et al. 2003; Mutschke 2003; White and Smyth 2003) and topic-based approach. The kernel of the topology-approach is to construct a graph on the basis of real-world entities and their interactions by using keywords-based matching. According to information flow between nodes in the graph, this approach focuses on link analysis to calculate expertise rank in the social network. By contrast, the topic-based approach captures the semantics-based information embodied in the documents, such as papers, blogs, discussion transcripts, and emails, produced by the network entities to estimate the probability that a candidate could be an expert on a certain topic.

Regarding the topology-based approach, the classical Webpage ranking algorithms PageRank (Brin and Page 1998) and HITS (Kleinberg 1999) are adopted to estimate the expertise of users by analyzing users' interactions, such as email conversations (Dom et al. 2003). Zhang et al. (2007) proposed an ExpertiseRank algorithm based on PageRank to generate a measure that not only considers how many other people one helped, but also whom he/she helped. Hilltop is another method by following two steps: expert page searching and target pages ordering (Bharat and Mihaila 2000). Schall (2012) presented a ranking model DSARank based on interaction metrics and advanced context-aware ranking techniques that make use of contextual-link information.

Furthermore, Kong presented a tweet-centric approach for topic-specific author ranking in Micro-Blog. In addition to traditional reply to relationship, the approach considers the influence of users according to their posting tweets on a certain topic. Nevertheless, the topic-specific tweets were simply extracted through keyword matching (Kong and Feng 2011). The advantage of this approach is that the link between the nodes reflects the real-world connection, and it was proved effective in applications such as Q/A forum and paper citation networks. Yet, this approach ignores the interested topics shared by members. As a result, each expertise network contains several topics, which mixes the meanings of the community (Ding 2011).

As for the topic-based approach, Li and Tang (2008) proposed a unified model that temporal information is modeled in a forward-and-backward propagation process in the random walk for experts finding. An expert search framework was also presented that divides expert search problem into three dimensions: ranking, quoted analysis, and topic maps search (Tang et al. 2011). Balog et al. (2009) proposed a language modeling framework for expert finding. Zhang et al. (2008) presented a mixture model for expert finding based on pLSA. Latent Dirichlet Allocation (LDA) is an unsupervised machine learning topic model, which is widely used and modified for expert finding (Daud et al. 2009, 2010; Rosen et al. 2004). The advantage of topic-based approach is that expert candidates discovered in one network have common interested topics. Nevertheless, the relationship between members only denotes the similarity in terms of produced artifacts but not real-world connection.

Therefore, neither social link analysis nor textual information alone is sufficient for discovering meaningful expertise network in online communities (Li et al. 2012). Based on computing of the similarity between document and concept in knowledge map for message topic recognition, Li et al. proposed to combine reply to relationship and topic similarity for special interest groups discovery within forums (Li et al. 2009). Zhao et al. (2012) proposed a topic-oriented community detection approach which combines both social objects clustering and link analysis. They presented a modified K-Means algorithm for object clustering, and then members involved in the social objects are partitioned into topical clusters.

4.2.2 Sentiment Analysis of Web Documents

Text sentiment analysis has become a flourishing frontier in the text mining community. The sentimental analysis can be divided into three tasks: sentiment extraction, sentiment classification, and sentiment retrieval and summarization (Zhao et al. 2010). Sentiment extraction is the basic task to extract the meaningful units, providing structured text for more applications. Based on the sentiment extraction, sentiment classification includes two kinds of classification forms, i.e., binary sentiment classification (positive and negative polarity) and multi-class sentiment classification (e.g., strong positive, positive, neutral, negative, and strong negative) (Tang et al. 2009).

Initial sentiment classification is based on the benchmark words with strong positive and negative polarity (Hatzivassiloglou and McKeown 1997). As there are more and more Web documents, sentiment classification processes the text in different language grains, including the word level, phrase level, sentence level, document level, and multi-documents level (Huang and Zhao 2008). Most work has been done on this type of task. Sentiment retrieval and summarization can be regarded as the direct interface with users, which focuses on retrieval and summarization based on the analyzing results of the former two types of tasks.

Existing sentiment analysis approaches fall into two categories: (1) semantic orientation-based approaches (Ahmad and Almas 2005; Du et al. 2009; Turney 2001; Turney and Littman 2003; Turney 2002; Xu et al. 2007, 2008; Yuan 2003) which make use of part-of-speech (POS) tagging of words and sentiment lexicons to determine the sentiment. Liu and Li (2002) proposed to establish a believable vocabulary on semantic knowledge named HowNet, and then get the sentiment polarity of words through comparison with the similarity between the words. Zhu (Zhu et al. 2006) succeeded in judging semantic orientation of Chinese online reviews based on the HowNet lexicon. (2) Machine learning-based approaches (Chaovalit and Zhou 2005; Hu et al. 2007; Tang et al. 2007; Turney 2002; Xu et al. 2007, 2008) which make use of different machine learning algorithms for sentiment classification at various levels. Whitelaw used support vector machine (SVM) to classify the sentiment orientation of movie reviews (Whitelaw et al. 2005). Lin proposed a learning method based on statistical model, which obtained opinions reflected in the text through analyzing words (Lin et al. 2006). Liao et al. (2009) made use of the probability statistical model to retrieval blog opinions. Chen (2011) proposes a new approach for sentiment classification by incorporating generated lexicons with the SVM algorithm.

User-generated reviews on the Web reflect users' sentiment about products, services, and social events. Yet, the existing researches mostly focused on the sentiment classification of the product and service reviews in document level. Only a few researchers pay attention to the sentiment analysis of the social reviews. Fu et al. (2013) adopt the similarity calculation to measure the word's sentiment orientation.

4.3 Expert Finding Approach

4.3.1 Overview of the Proposed Approach

Social context is the kernel to identifying expertise in diverse online learning communities. Compared with traditional contexts that are defined based on textual information, social context need incorporate various dynamic social relationships, such as the replying relationships between documents, follower–followee relationships between users. Figure 4.1 shows the social context. As the figure illustrates,

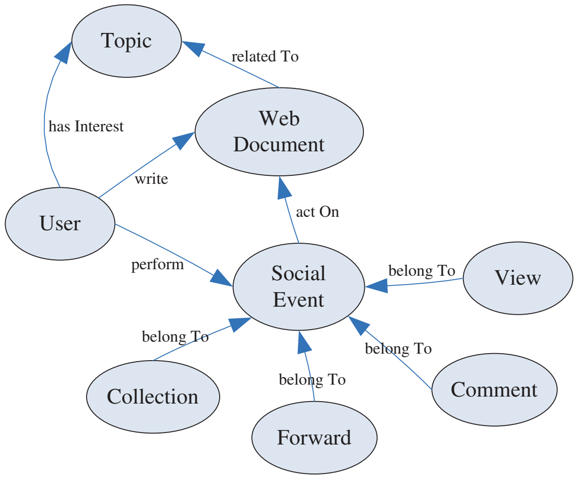

Fig. 4.1 Social context

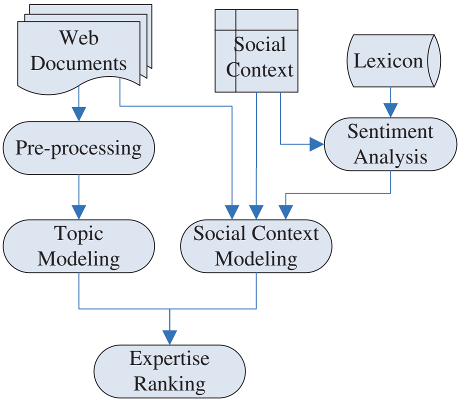

Fig. 4.2 Overview of the proposed approach

regarding one specific topic, the social context comprises of Web document information and social event information.

Definition 1 (*Social context*) Given a Web document d, its social context is defined as (U_d, E_d, M_d), where E_d denotes the social events performed by users U_d on d in an online learning community, M_d is a set of comments on d written by users U_d in an online learning community. $M_d = \{PM_d, NM_d, OM_d\}$, where PM_d is a set of comments with positive attitude toward d, OM_d is a set of comments with negative attitude toward d, NM_d is a set of comments with neutral attitude toward d.

Definition 2 (*Social event*) Social event $E_d = \{F_d, V_d, L_d\}$is defined as a set of actions performed on document d, where F_d, V_d, L_d denotes a set of forward, view, collection actions on d performed by users U_d, respectively.

Figure 4.2 illustrates the overview of the proposed approach. As the figure shows, the Web documents are firstly preprocessed for topic modeling.

The comments of the Web documents are analyzed to find out their sentimental attitude based on the lexicon. The social context information and sentiment analysis result along with the Web documents are then used for social context modeling. Following that, an algorithm is performed to compute user expertise for ranking.

4.3.2 Sentiment Analysis Based on HowNet Lexicon

HowNet lexicon (http://www.keenage.com/html/e_index.html) is a general knowledge base of Chinese and English words. Herein, we adopt the sliding window (Fu et al. 2013) to calculate the sentiment orientation based on HowNet lexicon. The basic idea is as follows: the orientation of the sentiment words in the sliding window is calculated through some positive and negative benchmark words, and then the orientation of the sliding window is summarized from its sentiment words. So, the key problem is how to determine the sentiment polarity of a sentiment word. We solve this problem by applying HowNet lexicon.

HowNet provides three types of lexicon: (1) positive and negative words, the polarities of positive words are set as +1 and that of negative words are set as −1; (2) privative words; (3) degree adverbs whose degree is set between −0.5 and 2.5, and thus such type of words provided in the HowNet are divided into six degrees. Additionally, we construct a lexicon of dynamic sentiment words (DSW) based on the domain, which comprises DSW and partner words (PW). DSW is divided into Forward DSW (e.g., high) and Backward DSW (e.g., low). As well, PW is divided into Forward PW (e.g., quality) and Backward PW (e.g., gas consumption). Given a dynamic sentiment word dw and partner word dp, its sentiment orientation SO(phrase) will be calculated as Eq. (4.1), where FDSW and BDSW denotes the forward-and-backward DSW, respectively, FPW and BPW denotes the forward-and-backward PW, respectively.

$$
\begin{aligned}
\mathrm{SO(dw)} &= \begin{cases} 1, & \text{if dw} \in \mathrm{FDSW} \\ -1, & \text{if dw} \in \mathrm{BDSW} \end{cases} \\
\mathrm{SO(dp)} &= \begin{cases} 1, & \text{if dp} \in \mathrm{FPW} \\ -1, & \text{if dp} \in \mathrm{BPW} \end{cases} \\
\mathrm{SO(phrase)} &= \mathrm{SO(dw)} * \mathrm{SO(dp)}
\end{aligned}
\tag{4.1}
$$

Given a review, the first step is to locate the matching benchmark sentiment words in the lexicons. Then take the matching words as focus, check the number of privative words in the forward sliding window and the degree adverbs in the sliding window (forward and backward). If the number of privative words is odd or even, then the polarities of sentiment words are set as −1 or +1. So, the orientation of sentiment words is to multiply the values of privative words, degree adverbs, and sentiment words. Finally, the sentiment orientation of a sliding window SO(window) can be calculated by the average sum of those orientation of its sentiment words.

$$\mathrm{SO(window)} = \sum_{i=1}^{n} \mathrm{SO}_i(\mathrm{phrase}) \tag{4.2}$$

Obviously, if SO(window) > 0, the sentiment polarity of the review is positive; if SO(window) < 0, that means the sentiment polarity of the review is negative; otherwise SO(window) = 0 implies a neutral sentiment polarity of the review.

4.3.3 Topic-Specific Expertise Ranking Based on Social Context

The crucial issue for expertise network discovery is to find users with high expertise on a specific topic. Our approach performs topic-specific user ranking based on two assumptions. The first is that the more documents (e.g., blogs, discussion transcripts) a user posted or commented on a specific topic, the more interest he/she has on this topic. The second is that a document commented or re-tweeted, forwarded by other influential users (i.e., users with high expertise) on the topic appears to have better quality than those commented, re-tweeted or forwarded by less influential users.

LDA (Blei et al. 2003) is a state-of-the-art topic model used to model documents by creating a latent topic layer between them. It is a Bayesian network that generates a document using a mixture of topics. In our study, we adopt LDA to obtain the document-topic matrix. Afterward, Eq. (4.3) is used to select the topic with the highest probability to which the document belongs, denoted as $P_z(d)$.

$$P_z(d) = \mathrm{Max}\{P(d|z)|z \in Z\} \tag{4.3}$$

where d indicates a document, z denotes a specific topic, and Z denotes the set of discovered topics.

- Reliability computing to reflect documents' quality

In online communities, users mainly play two roles, namely author and viewer of documents. The quality score of a document is subject to the feedback behaviors of its viewers (e.g., comment, forward), as well as the expertise scores of its linked users who may also be viewers or authors of other documents. Equation (4.4) is used to compute the reliability of a document on a certain topic z.

$$\begin{aligned}\mathrm{Rel}_{n+1}(z,d) = w_1 * \sum\nolimits_{i\in E_d} \left(\frac{\sum_{a\in U_d} (\mathrm{Exp}_n(z,a) * \mathrm{Num}_{a,i})}{\sum_{d\in D_z} \left(\sum_{a\in U_d} \left(\mathrm{Exp}_n(z,a) * \mathrm{Num}_{a,i} \right) \right)} \right) \\ + w_2 * \left(1 - \frac{|OM_d|}{|M_d|} \right)\end{aligned} \tag{4.4}$$

where E_d denotes the users' feedback behaviors on the document, and herein we consider actions including forward, view and collection. U_d denotes the set of users who perform the ith action on the document d. D_z denotes the set of documents that belong to the topic z. $\mathrm{Num}_{a,i}$ denotes the number of ith action performed by the author a on the document d. M_d is a set of comments on the document d, OM_d is a set of comments with negative attitude toward d, w_1 and w_2 are the adjustable parameters.

- Users' topic-focus activity measurement

It is common that users either post documents or provide feedbacks on documents in online communities. Accordingly, we define (4.5) and (4.6) to compute $\mathrm{PosAct}_z(a)$ and $\mathrm{BehAct}_z(a)$ of a user on topic z, respectively. The two indicators reflect users' contributions and interests on a specific topic. Then, Eq. (4.5) is defined to combine the two indicators for computing the topic-focus degrees of users. The greater the value of $\mathrm{Act}_z(a)$ is, the more the user focuses on the topic z.

$$\mathrm{PosAct}_s(a) = \frac{|D_{z,a}|}{\sum_{z=1}^{|Z|} |D_{z,a}|} \tag{4.5}$$

where $D_{z,a}$ denotes the set of documents posted by the user a on the topic z.

$$\mathrm{BehAct}_z(a) = \sum_{i \in E_d} \left(w_j^* \frac{\mathrm{Num}_{z,a,i}}{\sum_{z=1}^{|Z|} \mathrm{Num}_{z,a,i}} \right), \quad d \in D_z \tag{4.6}$$

where $\mathrm{Num}_{z,a,i}$ denotes the number of ith action performed by the author a on documents belonging to the topic z. To combine the diverse actions performed by users, we define (4.7) to measure user's topic-focus activity.

$$\mathrm{Act}_z(a) = \alpha^* \mathrm{PosAct}_z(a) + \beta^* \mathrm{BehAct}_z(a) \tag{4.7}$$

where α and β are the adjustable parameters.

- Topic-specific expertise ranking

According to the above-mentioned assumptions, we define (4.8) to compute the topic-specific expertise of an author through combination of document's reliability, documents' topic-focus degree, and user's activity.

$$\mathrm{Exp}_{n+1}(z,a) = \begin{cases} \frac{\Sigma_{d \in D_{z,a}} (P_z(d) * \mathrm{Rel}_{n+1}(z,d))}{|D_{z,a}|} * \mathrm{Act}_z(a), & |D_{z,a}| > 0 \\ \mathrm{Min}\{\mathrm{Exp}_n(z,a) | a \in A\}, & |D_{z,a}| > 0 \end{cases} \tag{4.8}$$

where A denotes the set of users. $|D_{z,a}|$ denotes the number of documents posted by a user on topic z, and $|D_{z,a}| = 0$ means that the user only give feedbacks to documents but not post any document. In such case, we believe that the expertise rank of such users is minimal compared to other users who post documents.

Note that computation of Exp(*z,a*) in (4.8) is an iteration process, due to the observation that the quality of the original documents posted by a user is determined by the expertise of users who give feedback. When the iteration process is completed, Exp(*z,a*) is used as the final ranking score of the user *a* on the topic *z*. Algorithm 1 shows the core pseudocode of the approach. The initial value of Exp(*z,a*) for each user is assigned 0.5.

```
Algorithm1. ExpRank {
//Given a topic-document list P_z(d), this algorithm is to obtain the expertise ranking of users on each topic.
for each user a∈U do
        Exp_0(z,a)←0.5
end for
n ← 0; //iteration indicator
for each topic z∈Z
    repeat
        compute  PosAct_z(a)
        compute  BehAct_z(a)
        compute  Act_z(a)
        for each document d∈D_z  do
            compute Rel_{n+1}(z,d) for each behavior i
        end for
        for each user a∈U
            if  |D_{z,a}|>0  do
                Exp_{n+1}(z,a) ← (Σ_{d∈D_{z,a}} P_z(d) * Rel_{n+1}(z,d)) / |D_{z,a}| * Act_z(a)
            else  do
                Exp_{n+1}(z,a) ← Min{Exp_n(z,a)|a∈U}
        end for
        n←n+1
    until Exp(a,z) convergence
    E_z← Rank(Exp(a,z))
end for
return expertise list E_z for z∈Z
}
```

Moreover, two indicators are used to scale experts' performance. Longevity describes an expert's persistence, which indicates the user's concentration stability on a topic in temporal terms. Centrality describes the topic-focus degree of experts.

- Longevity

Longevity is computed based on user behaviors, such as the number of postings, posting duration, and when he/she first delivered a document on a specified topic *z*. We sum up the number of documents posted by a user per month and calculate the standard deviation of the sum values, denoted as PostSD(*a*).

$$\mathrm{Lon}(a) = \left(LT_{z,a} - FT_{z,a}\right) \cdot \mathrm{PostAVG}(a) / \left(\left(CT_{z,a} - LT_{z,a} + 1\right) \cdot \mathrm{PostSD}(a)\right) \quad (4.9)$$

where $LT_{z,a}$ denotes the date of the most recent post, $FT_{z,a}$ denotes the date of the first post, PostAVG(a) denotes average number of posts, and PostSD(a) denotes standard deviation of number of posts. $CT_{z,a}$ denotes the collected data date.

- Centrality

$$\mathrm{Cen}(a) = \delta \cdot |D_{z,a}| / \sum_{z=1}^{|Z|} |D_{z,a}| + (1 - \delta) \cdot \mathrm{ReplyNum}_z / \sum_{z=1}^{|Z|} \mathrm{ReplyNum}_z \quad (4.10)$$

where $|D_{z,a}|$ denotes the number of documents posted by user a on topic z $\mathrm{ReplyNum}_z$ denotes the number of replies of user a on topic z. δ is an adjustment factor with a value range of (0, 1).

4.4 Experiment and Evaluation

In this section, we present the details of experiments on real datasets to evaluate the performance of our approach. We applied our approach to one real dataset and identified experts in online learning communities.

4.4.1 Data Collection

We chose one real dataset from Sina, which is one of the largest portal Websites with more than three hundred million users and provides the most popular blog in China, to investigate whether our approach is effective in identifying experts in online learning communities.

We crawled blogs and their properties such as publication time, author, replies, from Sina educational blog to launch our research. This dataset contained educational blogs from September 2008 to August 2011. We extracted 2,269 blogs, 738 authors who posted blogs and 28,313 users who only gave feedback without publishing any blogs. We discarded those replies where authors replied to themselves. We treated those people who published documents as authors even if they did not receive any replies to their postings.

4.4.2 Experiment Process and Results

First, we preprocessed the datasets by (a) splitting Chinese words using open-source software ICTCLAS (ICTCLAS, n.d.), (b) removing stop-words, punctuations, and numbers, (c) representing each document via vector space model for similarity calculation, (d) document transforming by integrating all the documents into a single document (where each document corresponds to a line represented as words) for input to LDA tool (MAchine, n.d.). Next, we adopted LDA to obtain the topic-word matrix and the document-topic matrix. As LDA requires pre-defining the number of topics, we conducted several tests by adjusting the number of topics and estimated topic models by using human judgment for obtaining meaningful results. The topics were set at seven for educational blog. The hyper parameters α and β of LDA model were set at 50/topic number and 0.01, respectively. The final result for the blog dataset is illustrated in Table 4.1. As shown in Table 4.1, each of the seven topics for the educational blogs is associated with six words that have the highest probability of occurrence, which provides a meaningful description of each topic.

On the basis of the discovered topics, the algorithm *ExpRank* was performed to obtain the expertise ranking for users on each topic. To simplify the algorithmic time complexity, we omitted those users who posting fewer than 2 times. Table 4.2 illustrates the top 10 experts with expertise probability on three hot topics. For example, user #248 has the highest expertise ranking on the topic "school education."

Table 4.1 The seven topics extracted from the blog dataset

Topics of educational blog	Top six words for each topic
Topic 0: school education	Student, teacher, study, school, education, problem
Topic 1: quality education	Society, life, work, spirit, China, leadership
Topic 2: English study	New oriental, Minhong Yu, world, problem, requirement, English
Topic 3: educational problems	Education, China, college, development, talents, society
Topic 4: college entrance recruiting	College entrance exam, recruit, matriculate, student, score, universities
Topic 5: college ranking	College, China, 2011, major, ranking, undergraduate
Topic 6: family education	Children, parent, educate, family, problem, psychology

Table 4.2 Top 10 experts on three hot topics

Topic 0		Topic 1		Topic 6	
School education		Quality education		Family education	
ExpertID	Prob.	ExpertID	Prob.	ExpertID	Prob.
248	0.00210	609	0.00339	503	0.00356
511	0.00192	197	0.00226	693	0.00244
13	0.00185	684	0.00147	154	0.00175
329	0.00101	589	0.00145	428	0.00171
154	0.00064	657	0.00129	620	0.00138
48	0.00052	333	0.00125	511	0.00135
394	0.00045	523	0.00097	481	0.00071
68	0.00036	236	0.00089	94	0.00069
37	0.00033	137	0.00081	612	0.00064
432	0.00028	78	0.00076	347	0.00062

4.4.3 Evaluation

As there were no explicit user-supplied expertise ranking data in the educational blog, we used human raters to generate a "gold standard" for comparison purposes. Because it was not possible for human raters to rank a large number of these users, we randomly selected 100 users as a comparison sample. Although the ranking algorithm produced continuous values that can potentially differentiate between all users, it is very difficult for humans to sort dozens of users into a ranked list. To evaluate a user's expertise level, raters must browse all documents (i.e., blogs or transcripts) he/she posted and others' feedback. It is also difficult to compare two users with similar performance.

To address this problem, Zhang et al. (2007) proposed letting raters categorize the users into five expertise levels instead of creating a complete ranked list based on observation and the results of a pilot rating set. Likewise, we selected three topics and invited two educational experts to rate the users' ranks into five levels.

After each human rater submitted his/her ratings, we tested the reliability of the raters by looking at their inter-rater correlation. The Kendall's tau_b distance between the two human raters was 0.835, and the Spearman's rho correlation coefficient was 0.877 ($p < 0.01$), which is sufficiently high.

To measure the performance for the automatic algorithm, we summed the ratings from the two raters into a standard human rating. Next, we performed significant correlation analysis between the algorithm and human ratings on different topics. Figure 4.3 shows the statistical correlations between our approach and the human ratings of the 100 users on three topics. As the figure shows, *ExpRank* algorithm gives a relatively high correlation to the human-assigned ratings, which implies that our approach effectively ranks opinion leaders in online learning communities.

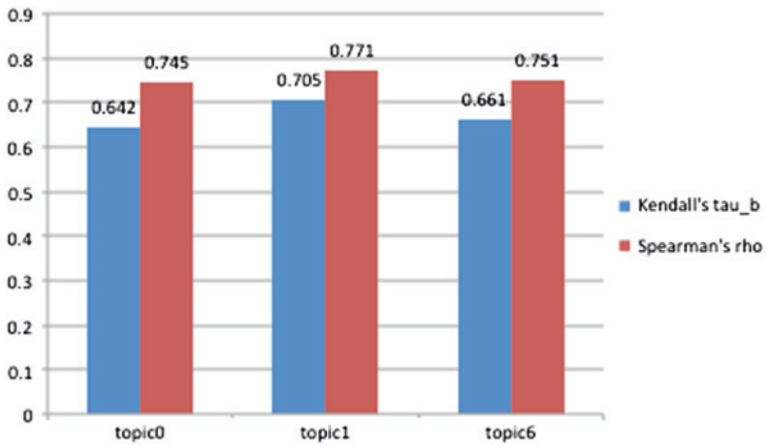

Fig. 4.3 The performance of the algorithm in two statistical metrics

4.5 Discussion

Most of the previous work related to expertise finding has focused on the Web-based research communities. The motivation is to find a person with topic related expertise to automatically fulfill different recommendation tasks by creating knowledge bases of researchers that would support the finding of appropriate collaborators for the project, selecting experts for consultation on some specific research topics, matching reviewers with academic papers, finalizing program committee members, and inviting keynote speakers for the conferences (Daud et al. 2010).

Cao et al. (2005) proposed a two-stage language model that combines a co-occurrence model to retrieve documents related to a given query, and a relevance model for expert search in those documents. Liu et al. (2005) studied a weighted directed co-author network and proposed AuthorRank algorithm for ranking authors. Mimno et al. (Mimno and McCallum 2007) proposed expertise modeling for matching papers with reviewers without considering conferences information. Tang et al. (2008) focuses on extracting and mining of academic social networks by considering the combined influence of conferences and authors for expertise search. In academic settings, the documents are academic papers and users are authors of the research papers, the journals or the conferences influence in which the papers published as well as citation can be very important factors to indicate the quality of papers. So, previous studies on expert finding focus on different approaches of topic modeling and analyzing network of co-author and citation in the academic field to find the "true" experts with greater professionalism and knowledge about specific subject.

As more and more people join different online learning communities for communication, knowledge sharing, and consultation, this study aims to explore the identification of experts in such open, free, and diverse communities. Finding experts in online learning communities is relatively informal recognition by people such as co-workers, acquaintances, or community members. Compared with normative academic papers, documents in online learning communities, such as blogs and forums, exhibit more informal oral language. User network is constituted via their behaviors such as view, comment, and forward. So, this study proposes to analyze social context of Web documents for expert finding. In addition to the content posted by the users, we consider user behavior and sentiment orientation to identify experts on different topics within a community. The performances of experts were further examined in terms of longevity and centrality, which can imply the characteristics of experts. We found that the experts are topic-specific; that is, a person who is an expert in one field may be a follower in another field.

With the popularization of online learning communities for informal learning such as blogs and forums, new massive open online courses (MOOCs) are attracting thousands of learners. MOOC is not a gathering but rather a way of connecting distributed learners as a learning community across a common topic or field of discourse (MOOC, n.d.). In such online learning communities with a large number of participants, experts are considered to be indispensable in making the evolution of learning communities steadier and learning more effective. So, two implications can be derived for educational scenario. First, the fundamental impediment to promotion of online learning communities as effective e-learning facilities lies in the large number of produced artifacts and learners with loosely tied social connections who have the ability to respond freely and informally. The possible solution could be to identify and recommend experts with high expertise and influence to make an online learning community interconnected and cohesive. Second, for the learners who would like to learn about a specific topic, it is reasonable to follow experts with high centrality or longevity for providing focused and in-depth information or advice.

One limitation of our approach is that we cannot determine the originality of documents in open and free learning communities if the author did not mark it. Yet in the research communities, this problem can be avoided because of the strict plagiarism checking. The second weakness is that we ignore the posting forwarding activity since its details are inaccessible in our experiment. Including this factor in the computational experiment would have a more convincing result. Moreover, the results are limited by the nature of the sample, since the datasets included only Chinese language content. Further studies are needed to ascertain the applicability of the results for other languages.

4.6 Conclusion

In this chapter, we explore analyzing social context to identify expertise in online learning communities. After preprocessing the datasets, document clustering was performed using the LDA approach, where the authors of a certain document cluster constituted a topic-specific network. Then, an algorithm is proposed to compute users' expertise rank based on documents' topic-focus degree, quality of documents, and users' activities. In addition, two other factors, namely longevity and centrality, were used to measure expert performance. Empirical study was conducted based on a real dataset from Sina educational blogs in China. The results and comparison show that our approach is feasible and effective to find meaningful expertise.

To systematically combine the content analysis and social behaviors represents a new and interesting direction for social networks. Our approach has potential applications in the context of learning. The experts identified by our approach are meaningful because they are topic-specific, which means they can facilitate knowledge sharing on specific topics and enhance the efficiency of collaborative learning. In practical applications, customized recommendation systems could be designed to recommend suitable experts to cater for learner's individual learning demands.

Acknowledgments This research work is supported by National Natural Science Foundation of China (NSFC: 61075048). We would like to thank Dr. Yonghe Zhang and Mr. Feipeng Sun for their work on data analysis.

References

Ackerman, M.S., & McDonald, D.W. (1996). Answer garden 2: Merging organizational memory with collaborative help. In *Proceedings of CSCW '96* (pp. 97–105). Boston, MA: ACM Press.

Ackerman, M. S., Wulf, V., & Pipek, V. (Eds.). (2003). *Sharing expertise: Beyond knowledge management*. Cambridge: MIT Press.

Ahmad, K., & Almas, Y. (2005). Visualising sentiments in financial texts? *Proceedings of the Ninth International Conference on Information Visualisation* (pp. 363–368).

Balog, K., Azzopard, L., & Rijke, M. (2009). A language modeling framework for expert finding. *Information Processing and Management, 45*, 1–19.

Bharat, K. & Mihaila, G. (2000). Hilltop: A search engine based on expert documents. *Poster Proceedings of WWW9* (pp. 72–73).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research,3*, 993–1022.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Proceedings of the 7th International Conference on World Wide Web (WWW '98)* (pp. 107–117), Brisbane, Australia.

Cao, Y., Liu, J., Bao, S., & Li, H. (2005). Research on expert search at enterprise track of TREC2005. *Proceedings of TREC2005*.

Chaovalit, P. & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classi fication approaches. *Proceedings of the 38th Hawaii International Conference on System Sciences 2005*.

Chen, B. (2011). Topic oriented evolution and sentiment analysis. *Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, UMI 3471730*.

Daud, A., Li, J., Zhou, L., & Muhammad, F. (2009). Exploiting temporal authors interests via temporal-author-topic modeling. *ADMA 2009, LNAI 5678* (pp. 435–443).

Daud, A., Li, J., Zhou, L., & Muhammad, F. (2010). Temporal expert finding through generalized time topic modeling. *Knowledge-Based Systems, 23*, 615–625.

Ding, Y. (2011). Community detection: Topological versus topical. *Journal of Informetrics,5*(4), 498–514. doi:10.1016/j.joi.2011.02.006.

Dom, B., Eiron, I., Cozzi, A., & Zhang, Y. (2003). Graph-based ranking algorithms for e-mail expertise analysis. *DMKD 2003* (pp. 42–48). ACM: New York, USA.

Du, W., Tan, S., Yun, X., & Cheng, X. (2009). A new method to compute semantic orientation. *Journal of Computer Research and Development, 46*, 1713–1720.

Erten, C., Harding, P.J., Kobourov, S.G., Wampler, K., & Yee, G. (2003). Exploring the computing literature using temporal graph visualization. Technical Report, *Proceeding of SPIE 5295, Visualization and Data Analysis 2004*, 45 (June 4, 2004), (pp. 45–56). doi:10.1117/12.539245, http://dx.doi.org/10.1117/12.539245

Foner, L.N. (1997). Yenta: A multi-agent, referral-based matchmaking system. In *Proceedings of Agents '97* (pp. 301–307). ACM Press: Marina del Rey, CA.

Fu, X.H., Liu, G., Guo, Y.Y., & Wang, Z.Q. (2013, January). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems* (Vol. 37, pp. 186–195). http://dx.doi.org/10.1016/j.knosys.2012.08.003

Hatzivassiloglou, V., & McKeown, K.R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the EACLv97* (pp. 174–181). ACL: Morristown. doi:10.3115/979617.979640.

Huang, X.J., & Zhao, J. (2008). Sentiment analysis for Chinese text. *Communications of CCF 2008*, *4*(2) (in Chinese with English abstract).

Hu, Y., Lu, R., Li, X., Duan, J., & Chen, Y. (2007). Research on language modeling based on sentiment classification of text. *Journal of Computer Research and Development,44*, 1469–1475.

ICTCLAS. (n.d.). Retrieved 26 Oct 2012 http://www.ictclas.org/

Kautz, H., Selman, B., & Shah, M. (1997). Referral web: Combining social networks and collaborative filtering. *Communications of the ACM, 40*(3), 63–65.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM, 46*(5), 604–632.

Kong, S. & Feng, L. (2011). A tweet-centric approach for topic-specific author ranking in micro-blog. ADMA 2011, Part I, LNAI 7120 (pp. 138–151), Beijing, China.

Krulwich, B., & Burkey, C. (1996). ContactFinder agent: Answering bulletin board questions with referrals. In *The 13th National Conference on Artificial Intelligence* (pp. 10–15), Portland, OR.

Lappas, T., Liu, K. & Terzi, E. (2011). A survey of algorithms and systems for expert location in social networks. In C.C. Aggarwal (Ed.), *Social network data analytics, Chap. 8* (pp. 215–241). doi:10.1007/978-1-4419-8462-3_8.

Liao, X., Cao, D., Fang, B., Xu, H., & Cheng, X. (2009). Research on blog opinion retrieval based on probabilistic inference model. *Journal of Computer Research and Development, 46*, 1530–1536.

Littlepage, G. E., & Mueller, A. L. (1997). Recognition and utilization of expertise in problem-solving groups: Expert characteristics and behavior. *Group Dynamics: Theory, Research, and Practice,1*, 324–328.

Lin, W.-H., Wilson, T., Wiebe, J., & Hauptmann, A. (2006). Which side are you on?: Identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Natural Language Learning, Publishing* (pp. 109–116), New York.

Liu, Q., & Li, S. (2002). Word semantic similarity computation based on HowNet. *Proceedings of the 3rd Symposium on Chinese Words Semantics* (Vol. 5).

Liu, X., Bollen, J., Nelson, M. L., & de Sompl, V. H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management, 41*(6), 1462–1480.

Li, Y., & Tang, J. (2008). Expertise search in a time-varying social network. In *Web-Age Information Management, 2008. WAIM'08. The Ninth International Conference on* (pp. 293–300). IEEE.

Li, Y., Dong, M., & Huang, R. (2009). Towards a semantic forum for active collaborative learning. *Journal of Educational Technology and Society, 12*(4), 71–86.

Li, Y., Ma, S., Zhang, Y., & Huang, R. (2012, July). Expertise network discovery via topic and link analysis in online communities. *Proceedings of the 12th International Conference on Advanced Learning Technologies* (pp. 311–315), Rome, Italy.

MAchine Learning for Language Toolkit (n.d.). Retrieved 26 Oct 2012 http://mallet.cs.umass.edu/download.php

Mimno, D., & McCallum, A. (2007). Expertise modeling for matching papers with reviewers. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 500–509).

MOOC (n.d.). Retrieved 26 Oct 2012 http://en.wikipedia.org/wiki/Massive_open_online_course

Mutschke, P. (2003). Mining networks and central entities in digital libraries: A graph theoretic approach applied to co-author networks. *Intelligent data analysis*, (pp. 155–166).

Rosen, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004, July). The author-topic model for authors and documents. *Proceedings of the 20th Conference on Uncertainty Inartificial Intelligence* (pp. 487–494), Banff, Canada.

Schall, D. (2012). Expertise ranking using activity and contextual link measures. *Data and Knowledge Engineering, 71*(1), 92–113.

Streeter, L., & Lochbaum, K. (1988). Who knows: A system based on automatic representation of semantic structure. In *Proceedings of RIAO* (pp. 380–388).

Tang, H., Tan, S., & Cheng, X. (2007). Research on sentiment classification of chinese reviews based on supervised machine learning techniques. *Journal of Chinese Information Processing, 21*(88–94), 108.

Tang, H. F., Tan, S. B., & Cheng, X. Q. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications., 36*(7), 10760–10773.

Tang, J., et al. (2011). Topic level expertise search over heterogeneous networks. *Machine Learning, 82*, 211–237. doi:10.1007/s10994-010-5212-9.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 990–998).

Turney, P.D. (2001). Mining the web for synonyms: PMI-IR versus LSA on TOEFL. *Proceedings of the Twelfth European Conference on Machine Learning* (pp. 491–502). Springer, Berlin.

Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems, 21*, 315–346.

Turney, P.D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 417–424). Publishing, Philadelphia, Pennsylvania.

Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 625–631). Publishing, Bremen, Germany.

White, S. & Smyth, P. (2003). Algorithms for estimating relative importance in networks. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 266–275).

Xu, D. J., Liao, S. S., & Li, Q. (2008). Combining empirical experimentation and modeling techniques: A design research approach for personalized mobile advertising applications. *Decision Support Systems, 44*(3), 710–724.

Xu, J., Ding, Y., & Wang, X. (2007). Sentiment classification for Chinese news using machine learning methods. *Journal of Chinese Information Processing, 21*, 95–100.

Yuan, S. (2003). A personalized and integrative comparison-shopping engine and its applications. *Decision Support Systems, 34*(2), 139–156.

Zhang, J., Tang, J., Liu, L., & Li, J. (2008). A mixture model for expert finding. In *Advances in knowledge discovery and data mining* (pp. 466–478). Springer, Berlin.

Zhang, J., Ackerman, M., & Adamic, L. (2007, May). Expertise networks in online communities: Structure and algorithms. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 221–230). ACM.

Zhao, Y., Qin, B. & Liu, T. (2010). Sentiment analysis. *Journal of Software*, *21*(8), 1834–1848. doi:10.3724/SP.J.1001.2010.03832. http://www.jos.org.cn/.1000-9825/3832.htm

Zhao, Z., Feng, S., Wang, Q., Huang, J. Z., Williams, G. J., & Fan, J. (2012). Topic oriented community detection through social objects and link analysis in social networks. *Knowledge-Based Systems, 26*, 164–173.

Zhu, Y., Min, J., Zhou, Y., Huang, J., & Wu, L. (2006). Semantic orientation computing based on HowNet. *Journal of Chinese Information Processing, 20*(1), 14–20.

Part II
Ambient Design

Chapter 5
Context-Aware Geography Field Trip with EagleEye: Teachers' First Experience

Morris S. Y. Jong

Abstract EagleEye is a global positioning system (GPS)-supported integrated educational system for empowering teachers and students respectively to facilitate and pursue context-aware outdoor exploratory learning in Geography field trip activities. We conducted a quantitative study (complemented with qualitative methods) to investigate 98 teachers' pedagogical perceptions of EagleEye and collect their feedback on its user-friendliness. Results showed that the teachers had positive perceptions towards this educational innovation, in terms of student-centredness, motivation, scaffolding, and user-friendliness; however, they had reservation about if EagleEye could better promote collaboration among students during a field trip activity. The findings shed light on our future work on implementing EagleEye in real teaching practice and provide insights for those who are engaging in developing, adopting, or appropriating mobile technologies to support learning and teaching.

Keywords Mobile learning · Context-awareness · Exploratory learning · Geography field trip · EagleEye · Teachers' perceptions

5.1 Research Motivation and Contribution

With the advocacy of learner-centric pedagogical paradigms in the twenty-first century, technological researchers in the field of education have been endeavouring to study how to design, implement, and evaluate mobile learning from various constructivist learning and teaching perspectives (e.g. Ogata et al. 2008; Pachler et al. 2013; Sharples 2002; Shih et al. 2010; Wake and Wasson 2011; Zurita and Baloian 2012). In this chapter, we will discuss our mobile learning initiative, *EagleEye*, for supporting context-aware outdoor field trip activities in Geography education.

M.S.Y. Jong (✉)
Department of Curriculum and Instruction and Centre for the Advancement of Information Technology in Education, The Chinese University of Hong Kong, Hong Kong, China

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_5

EagleEye is a global positioning system (GPS)-supported integrated educational system for empowering teachers and students respectively to facilitate and pursue exploratory learning in outdoor field trips. The aim of our development of this system is to mitigate four generic problems taking place commonly in conventional outdoor field trips: (i) *learning takes place in a teacher-centred manner*, (ii) *students' learning motivation cannot be sustained*, (iii) *scaffolds are not effective enough*, and (iv) *collaboration among students is weak* (see Sect. 5.3 for detailed discussion). Preliminarily, we obtained positive results in our previous pilot study on EagleEye with 38 Grade-9 students (Jong 2013a). After experiencing an EagleEye field trip, comparing with their prior conventional outdoor field trip experiences, the students perceived that EagleEye could provide them with a more desirable field trip experience in terms of *Student-centredness*, *Motivation*, *Scaffolding*, and *Collaboration*. However, since only two Geography teachers were involved in that study, we had yet to be able to give any conclusive remarks on teachers' feedback on EagleEye.

In 2013, we conducted a further study with 98 Geography teachers to evaluate EagleEye. This study aimed at investigating their perceptions of whether EagleEye could mitigate the four problems taking place in conventional outdoor field trips, and collecting their feedback on its user-friendliness. All of these teachers had participated in 5-h training on the use of this system before they provided their evaluative feedback in a questionnaire-based survey.

The objectives of this chapter are twofold. Firstly, it delineates our EagleEye initiative and the rationales behind our work. Secondly, it presents and discusses our study on teacher evaluation of EagleEye. We believe that the design of EagleEye and the research findings delineated in this chapter can provide insights for those who are working on the development, adoption or appropriation of mobile technologies, devices, and/or apps for supporting learning and teaching. We organize the rest of the chapter as follows. Section 5.2 is a review of the recent-related work in the domain. In Sect. 5.3, we will elucidate the rationales behind our development of EagleEye. Section 5.4 is a description of EagleEye. Sections 5.5 and 5.6 will delineate respectively the research design and findings of the present study. We will give our concluding remarks and discuss our future work in Sect. 5.7.

5.2 Literature Review

Experiential learning (Baker et al. 2002; Kolb 1984), which is one of the constructivist pedagogical paradigms being advocated in the twenty-first century education, emphasizes learning as the process whereby knowledge is created through the transformation of experience. One of the pedagogical approaches to implementing this learning paradigm is *field trips* (Douglass 2008; Frew 1999; Lenon and Cleves 2001).

Field trips can be categorized into *indoor field trips* and *outdoor field trips* (Nadelson and Jordan 2012). Class visits to museums, art/science centres, institutions, etc., are typical indoor field trip examples. A number of researchers (e.g. Shih et al. 2010; Nordmark and Milrad 2012) have endeavoured to harness mobile technology in transforming the learning taking place in traditional indoor field trips into more enquiry-oriented with the infusion of self-directed and/or collaborative elements. Class visits to country parks, forests, wetlands, villages, or other sites with specific natural, geographical, or heritage settings are typical outdoor field trip examples (Knapp and Barrie 2001). Our research focuses on the crossover between outdoor field trips and mobile learning. For writing convenience, the term "field trip(s)" refers to "outdoor field trip(s)" hereafter.

Field trips appear frequently in Geography, Ecology, Cultural, and even Language education for providing students with opportunities to pursue exploratory learning in real-life, real-world environments. In the recent years, a number of new mobile learning initiatives have adopted location-awareness technology, GPS (Ashbrook and Starner 2002; Shuler 2009) in particular, to create new forms of field trips. For example, Ogata et al.'s (2008) language-learning outside the classroom with handhelds (*LOCH*) is a PDA-based mobile learning system designed for supporting overseas students to learn the Japanese language and culture in real-life situations through accomplishing a number of exploratory tasks in a field trip. The tasks include interviewing local people, going to some specific places to gather information, buying some specific Japanese food, etc. During the field trip, the students can use their own PDA to jot down annotations, record interviews, and take pictures. After returning to the classroom, they will share and discuss their field trip artefacts with their peers. Nevertheless, the integration between the GPS's location-awareness feature and the exploratory learning opportunities offered in LOCH is weak. The main function of the GPS in the system is for students to know their own current geo-location to avoid getting lost during a field trip.

Wake and Wasson's (2011) *SILO* is a web-based development tool designed for facilitating students in groups to create phone-based location-aware games (then to be played by other groups in field trips) for learning the local history of Bergen. To create a SILO game, a group uses the game editor to tie a "storyline" to a geographical area on a digital map. Within the area, the group can set up so-called points of interest by clicking at various places on the map so as to specify the corresponding geo-locations defined by the GPS. At each point of interest, the group can embed text description of the place as well as clues for finding the next point of interest for arousing players' attention. When playing the game, the interface will show a marker displaying one's current position on the map, a track displaying the history of one's movement, and a bar displaying the icons of the places that the one has visited. Nevertheless, as underlined by Wake et al., the focus of SILO is on students' creative process of developing mobile games, rather than their actual field trip experiences gained through playing the games.

Zurita and Baloian (2012) have developed a prototype tablet-based system to implement their pedagogical proposition of "geo-collaborative learning with patterns" for facilitating students in field trips to "understand the patterns of neoclassical architecture found in the city" and "study the reasons of why certain patterns of trees appear more often in city parks" (p. 507). Before a field trip, with the system's teacher editor, a teacher can define various patterns with each annotated with a name and description over a digital map. He/she will then embed tasks (with guiding instructions) on the map for his/her students. These tasks usually involve data collection on the pattern instances through exploring a geographic area with upon a designated path. During the field trip, the students will work in groups with the system's client-side application run on their GPS-enabled tablets. When they find certain elements aligning with the patterns defined by their teacher, they will create instantiations of the patterns with text descriptions or sketches.

In Zurita and Baloian (2012) work, we can observe that they infused two important pedagogical elements into their mobile learning initiative. The first element is *teacher scaffolding* (Vygotsky 1978). Unlike many other mobile learning systems (developed solely for research purpose, containing fixed, unamendable contents), with the provision of an authoring tool (the teacher editor), Zurita et al.'s system allows teachers to create their own exploratory resources in accordance with their students' needs. However, the variety of exploratory scaffolds (to be embedded into the resources) offered in the teacher editor is narrow. The scaffolds are merely designed for their proposed pattern-based learning pedagogy. The transferability of the system to other kinds of field trips is weak. The second element is *teacher debriefing* (Lederman 1992). As emphasized by Zurita et al., a pattern-based learning field trip should end with a debriefing session for helping students conclude and reflect on what they have learned in the activity. In fact, teacher debriefing is regarded as one of the most important components of any kinds of field trips (Frew 1999; Lenon and Cleves 2001). Nonetheless, there is no specific after-the-field trip tool available in Zurita et al.'s system for better supporting teachers to carry out debriefing work.

5.3 Rationales Behind Our Work

Field trips are always one of most vital components in Geography education, because fieldwork is "the way in which our geographical knowledge of the world has been built up" (Frew 1999, p. 4). Conventionally, field trips in Geography education are organized and led by teachers (Douglass 2008). Before a field trip, a teacher will design a set of paper-based worksheets which aim at supporting his/her students in groups to observe, experience, and reflect, in accordance with some specific learning objectives. These worksheets usually contain "exploratory hints" (e.g. a number of closed-ended and short open-ended questions) for facilitating the students during the field trip to think about and respond to. At the end of the field trip, the students have to submit the teacher their "answers" to these questions in written format.

Conventional field trips, however, have a number of limitations. In our qualitative investigation (Jong et al. 2012; Jong 2013b) through in-depth interviews with 12 senior secondary students[1] and 12 Geography teachers,[2] we generalized the following four generic problems taking place frequently in conventional field trip activities:

[1] All of the students had several field trip-participation experiences (four to five times) when they were in junior secondary grades.

[2] All of the teachers had abundant field trip-facilitation experiences in Geography Education.

- *Problem 1*: *Learning takes place in a teacher-centred manner.* Owing to the consideration of cost-effectiveness, a field trip is usually organized in a "whole-class" basis. Commonly, the student-to-teacher ratio is large (two teachers facilitating a class of 40 students, or even more). A typical field trip setting is that teachers bring a whole class to the exploratory spots on a field trip site one by one in a designated (teacher-centred) order, without allowing students to plan and adjust their own exploratory route. This violates largely the original advocacy of the field trip approach to enabling learners to learn in a constructivist, student-centred fashion (Nadelson and Jordan 2012).
- *Problem 2*: *Students' learning motivation cannot be sustained.* At the beginning of a field trip, students' learning motivation is high in general (Hofstein and Rosenfeld 1996; Bell et al. 2009). They are happy and excited, because outdoor activities are relatively rare in comparison with everyday classroom activities. However, their motivation will decrease gradually during the field trip. A reason to this phenomenon is that they have no right to control the time being spent on each exploratory spot according to their own learning pace and interest. The time for staying at each spot is not determined by themselves, but controlled by their teachers. Furthermore, every time when a class of 40 students arrives at a single exploratory spot (usually a small area), it is too crowded for everyone to observe and experience the geographical context around and the happenings (e.g. social and cultural events) therein.
- *Problem 3*: *Scaffolds are not effective enough.* In a field trip, paper-based worksheets (composed of a number of closed-ended and short open-ended questions) are the primary learning support to students. These "hard scaffolds" (Brush and Saye 2002) are mainly text-based (or sometimes with images, such as, maps, pictures, etc.) presented in a static, boring manner. During the field trip, the students can only use text to respond to the questions by writing on the spaces aside or below the questions, or on the opposite sides of their worksheets. This paper-and-pencil approach does not appeal to today's youngsters who are, in Prensky's (2010) terms, "digital natives". They are more eager to have technological and multimedia tools for supporting their learning process.
- *Problem 4*: *Collaboration among students is weak.* In a field trip, although students are usually divided into groups and asked to pursue the exploration collaboratively, most of them only care about whether they can complete their worksheets in hand before the end of the activity. Instead of having much discussion with their groupmates, they spend a lot of time on copying the information from the field trip site for "entertaining" the questions on their worksheets. This kind of rote copying work, however, will never contribute to deep learning (Biggs and Moore 1993).

The aim of our development of EagleEye is to try to mitigate the above problems. The next section will spell out the details of EagleEye, and how it can support teachers and students respectively in facilitating and pursuing exploratory learning in field trip activities.

5.4 EagleEye

Adopting the paradigm of design-based research (Design-Based Research Collective 2003; Wang and Hannafin 2005), we create EagleEye—an integrated outdoor exploratory educational system for supporting field trip activities in Geography education. This is a second version of EagleEye with the provision of new types of exploratory scaffolds for facilitating outdoor exploratory learning. The design of these new scaffolds has been based on the results of the need analysis that we conducted with the two Geography teachers who participated in our previous pilot study (Jong 2013a) and the three Geography teachers in our project team.

EagleEye is composed of four core components, including the (i) *Location-aware Exploratory Resource Authoring Tool* (*LERAT*), (ii) *GPS-supported Exploratory Platform* (*GEP*), (iii) *Repository Server* (*RS*), and (iv) *Teacher Console* (*TC*). Figure 5.1 shows the interaction between the LERAT, GEP, TC, and RS respectively before, during, and after a field trip activity.

The **LERAT** is a piece of PC-based software for teachers to create *location-aware exploratory resources* run on GPS-enabled tablets. Figure 5.2 shows the LERAT's interface. A location-aware exploratory resource is a combination of a *map image* and a number of *location-based exploratory scaffolds*. When a teacher creates a location-aware exploratory resource, first of all, he/she needs to import a digital map file into the LERAT for designating the geo-area of the field trip site. Hence, the imported map becomes the backdrop of the resource (see Fig. 5.2). He/she then needs to conduct simple calibration by inputting the corresponding real-world latitude and longitude values[3] at the corners on the backdrop. After that, the teacher can start to set up, at each designated exploratory spot on the map, a location-based exploratory scaffold. Usually, each scaffold contains background information and guiding questions/tasks for hinting students to explore that spot. Various gadgets are available in the LERAT for the teacher to design the scaffolds with the use of multimedia elements (text, graphics, audio, or/and video) based on the pedagogical objectives of the field trip. The scaffolds can be in the form of an introductory briefing, multiple-choice or open-ended question, quantitative or qualitative data collection task, concept-map construction task, etc. The scaffolds created will then appear as circle-shaped "hotspots" (see the circles in Fig. 5.2). After developing the resource, the teacher will upload it to the RS.

The **GEP** is a tablet-based software application (i.e. an App). The current version is designed to run on Apple™ iPads, and is available for free download from the Apple™ App Store (see Fig. 5.3). Figure 5.4 shows the GEP's interface. Before a field trip, at a place (e.g. at school) with Internet access, each student group will connect to the RS to download the corresponding location-aware exploratory resource (created by their teacher) to their GPS-enabled tablet. At

[3] These values can also be obtained easily by right-clicking the corresponding locations on free online maps.

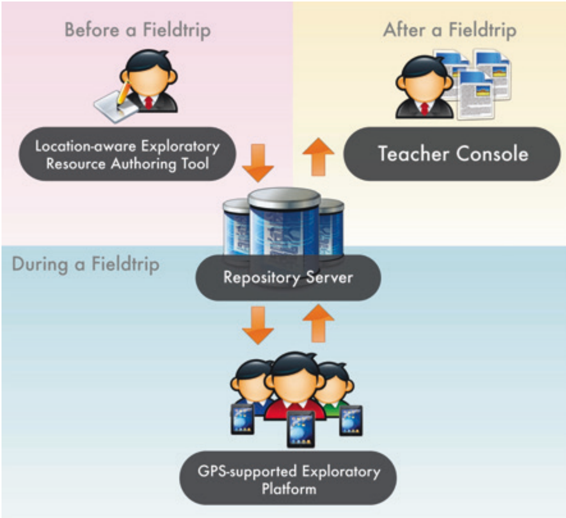

Fig. 5.1 Interactions between four components of EagleEye in a field trip activity

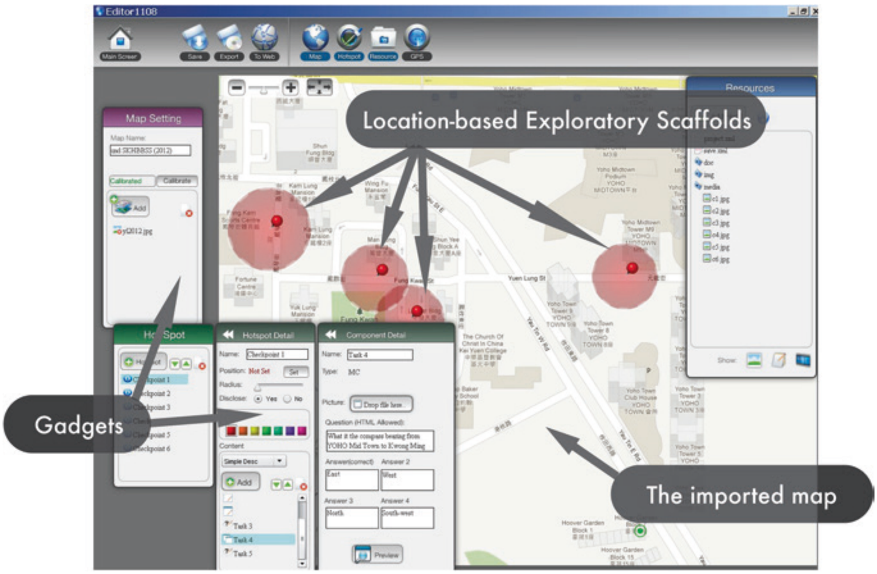

Fig. 5.2 Location-aware exploratory resource authoring tool (*LERAT*)

the beginning of the field trip, the groups will open the resource with the GEP. An "avatar" (see Fig. 5.4) will appear on the map to indicate their current geo-location in the real world. Based on the ongoing GPS signals received by the

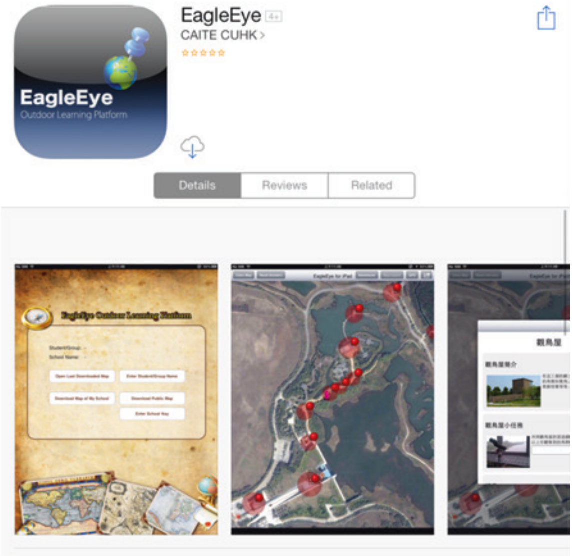

Fig. 5.3 EagleEye's GEP on Apple™ app store

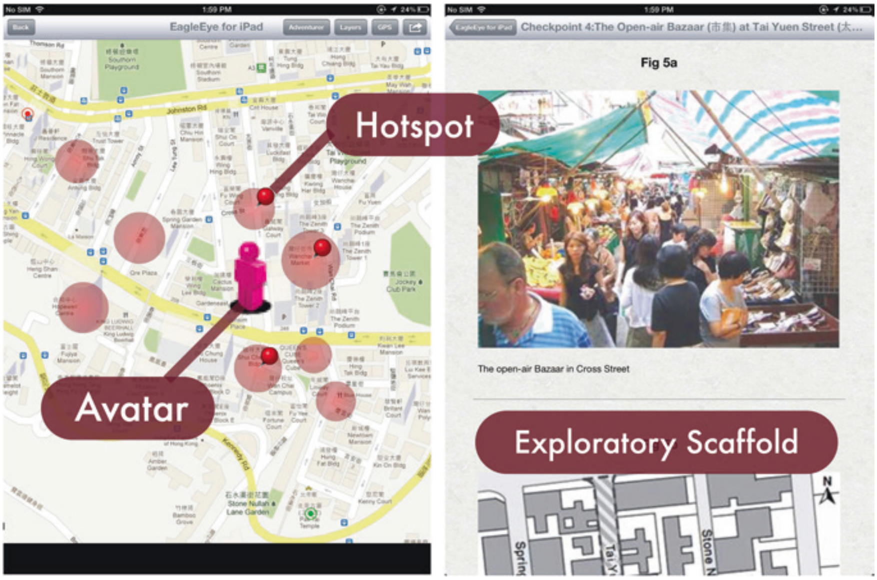

Fig. 5.4 GPS-supported exploratory platform (*GEP*)

GEP, the "hotspots" (i.e. location-based exploratory scaffolds) embedded in the resource will pop up automatically when the groups step in physically the corresponding geo-locations on the field trip site. The background information as well as guiding questions/tasks in these scaffolds will guide them to observe, experience, and reflect in the whole course of the field trip. Unlike conventional field trip activities where a whole-class crowds into each exploratory spot simultaneously, with the GEP the groups can plan and adjust their own exploratory route according to their own learning pace and interest. They can also determine themselves how long they want to spend on each spot. When exploring the spots, the groups will input their responses to the questions/tasks (presented in the scaffolds), and the GEP will save their responses automatically. No Internet access (neither Wi-Fi nor 3G) is needed during the field trip. After the field trip, the groups can upload their responses to the RS at a place with Internet access (e.g. after coming back to school).

The **RS** serves for three purposes. First, it stores location-aware exploratory resources created and uploaded by teachers. Second, it allows student groups before a field trip to download the corresponding resource to their GPS-enabled tablets. Third, it holds the responses (to the exploratory scaffolds) uploaded by student groups after a field trip for teachers' later retrieval.

The **TC** is a web-based platform connected to the RS. It aims at enabling teachers to retrieve student groups' responses to the exploratory scaffolds (on a location-aware exploratory resource) after a field trip for debriefing purpose. Figure 5.5 shows the TC's interface. It offers two viewing modes, namely *Group Display* and *Question Display*. The former is for displaying each group's responses to the exploratory scaffolds in an order with respect to their own exploratory route (the left side of Fig. 5.5). The latter is for comparing different groups' response to the same scaffold in a side-by-side fashion (the right side of Fig. 5.5).

Fig. 5.5 Teacher console (*TC*)

5.5 Research Design

The objectives of the present study were to (i) investigate teachers' perceptions of whether EagleEye could mitigate the four problems taking place in conventional field trips (as discussed in Sect. 5.3), and (ii) collect teachers' feedback on the user-friendliness of EagleEye. The research subjects were Hong Kong secondary Geography teachers who participated in our free induction training on EagleEye for supporting field trip activities.

5.5.1 Induction Training

We held six identical training events at our university on weekday evenings. Teachers were eligible to enrol the training only if (i) they were currently teaching Geography in secondary schools, and (ii) they had experiences in organizing and facilitating conventional field trips for their students. The maximum "class-size" of each event was 25, on a first-come-first-serve basis, through online enrolment. Four weeks before we commenced the training, we had sent the promotion flyer to all Hong Kong secondary schools through facsimile.

Each training event was composed of two sessions (namely, Session 1 and Session 2) which were respectively held on the same weekday (e.g. Monday) in two consecutive weeks. The duration of each session was 2.5 h, i.e. 150 min (excluding the recess time). Table 5.1 shows the rundown of each session. We (one instructor and two helpers) carried out the training mainly in a computer laboratory, except the field trip part in Session 1. Each participant was provided with an Apple™ iPad (pre-installed with the GEP) and an HP™ desktop computer (pre-installed with the LERAT). During the training, the helpers provided just-in-time support to the participants whenever necessary.

The field trip in Session 1 took place inside the campus of our University. The theme of the field trip was about exploring how the University preserves campus tree

Table 5.1 Training rundown

Session	Content
Session 1	Briefing on the agenda of the whole training (5 min)
	Elaboration on the pedagogical idea of EagleEye (20 min)
	Introduction to EagleEye (10 min)
	Demonstration of the GEP's operation (10 min)
	Briefing on the field trip (10 min)
	Pursuing the field trip with the GEP (75 min)
	Debriefing with the TC (20 min)
Session 2	Recap on Session 1 (10 min)
	Hands-on practice on the LERAT's operation (45 min)
	Hands-on practice on the TC's operation (15 min)
	Transforming conventional field trip materials into location-aware exploratory resources (60 min)
	Summary of the training and Q&A time (20 min)

and raises co-awareness among the staff and students in the University. During the field trip, the participants were divided randomly into groups of three or four in playing the role of students to pursue outdoor exploratory learning with the GEP. After the field trip, the participants came back to the computer laboratory and uploaded their responses to the exploratory scaffolds to the RS. Then, with the TC, we retrieved the responses and played the teacher role to debrief what they had observed and learned in the field trip, and to facilitate them to reflect on their own field trip experience.

In Session 2, we trained the participants the LERAT's operation, with step-by-step guidance on creating the location-aware exploratory resource that they had used in the field trip in Session 1. The participants also had hands-on practice on the TC's operation. After that, they were given time to transform their own conventional paper-based field trip material into a location-aware exploratory resource run on the GEP.

5.5.2 Data Collection

We adopted a quantitative research approach (with a questionnaire-based survey) to achieving the objectives of this study. The collection of the qualitative data was for the purpose of complementing the quantitative findings.

Quantitative Data. We administered the post-training questionnaire right after the completion of Session 2 in each training event. The questionnaire consisted of three parts. All items in the first two parts were in 5-point Likert scale (5: Strongly Agree, 4: Agree, 3: Neutral, 2: Disagree, to 1: Strongly Disagree). Part 1 of the questionnaire was a customized version of the instrument that we had developed for our previous pilot study (Jong 2013a). The original instrument was to study students' perceptions of their EagleEye-field trip experience in comparison with their past experiences in conventional field trips so as to from the learner perspective evaluate whether the EagleEye approach could mitigate the four problems discussed in Sect. 5.3. The instrument was composed of four subscales including Student-centredness, Motivation, Scaffolding, and Collaboration. Each subscale was constructed by two question items (i.e. totally eight items), articulating to one of the four problems. In the present study, we revised the subject and tense used in the eight items from (i) "I" to "My students" and (ii) past tense to future tense so as to from the teacher perspective evaluate whether the EagleEye approach could mitigate these four problems too. Table 5.5 (in Sect. 5.6.1) displays the eight items[4] and their corresponding subscales. Part 2 of the questionnaire contained three question items to gather the participants' feedback on the user-friendliness of EagleEye. Table 5.7 in Sect. 5.6.2 displays these three items. Part 3 of the questionnaire was to collect the participants' demographic data (including their age, years of teaching experience, and their schools' banding[5]).

[4] The items were presented in a random order in the questionnaire in the study.

[5] Secondary schools in Hong Kong are categorized into three academic bands based on their students' academic achievement; Band-1, Band-2, and Band-3 are respectively the top, middle, and bottom.

Qualitative Data. We collected the qualitative data through observations and participant–researcher conversations. Apart from conducing the training, we (the instructor and the 2 helpers) were also the observers in each training event. During the training, we also chatted a lot with the participants. After the completion of every event, we composed a journal for recording all major observations and conservations therein. The journals provided us with useful information to complement the survey results.

5.6 Findings and Discussion

There were 112 teachers participating in the six induction training events. We received 102 questionnaires in total; the return rate was 91.07 %. Four of the received questionnaires were excluded because the back-pages were omitted. Thus, the total number of questionnaires in our data analysis was 98. Tables 5.2, 5.3, and 5.4 display these 98 teachers' demographic information in groups, in terms of their age, teaching experience (in years), and school banding. The following subsections will present and discuss our findings with respect to the two research objectives spelt out in Sect. 5.5.

5.6.1 Mitigation of the Four Problems Occurring in Conventional Field Trips

Table 5.5 displays the subscales and descriptive statistics of the eight items in Part 1 of the questionnaire. Table 5.6 displays the reliability estimates and descriptive

Table 5.2 Age groups ($N = 98$)

Age	No. of teachers
35 or below	34
36–45	35
46 or over	29

Table 5.3 Teaching-experience groups ($N = 98$)

Teaching experience (in years)	No. of teachers
5 or less	32
6–10	35
11 or more	31

Table 5.4 School-banding groups ($N = 98$)

School banding	No. of teachers
Band 1	32
Band 2	31
Band 3	35

Table 5.5 Subscales and descriptive statistics (Part 1 of the questionnaire)

Comparing to my past field trip-facilitation experiences, I perceive that …

Subscale	Item	Mean	SD
Student-centredness	1. My students will participate more actively in an EagleEye field trip.	4.40	0.51
	2. My students will have better control on the process of exploration according to their learning pace and interest in an EagleEye field trip.	4.39	0.53
Motivation	3. My students will be more motivated in the process of exploration in an EagleEye field trip.	4.30	0.53
	4. My students will be more engaged in the process of exploration in an EagleEye field trip.	4.36	0.43
Scaffolding	5. My students will have better support in the process of exploration with the location-aware features of EagleEye *(the indication of geo-locations, and the just-in-time pop-up of hotspots)* in an EagleEye field trip.	4.23	0.57
	6. My students will have better support in the process of exploration with the multimedia and interactive features of EagleEye *(the hotspots, and the questions/tasks presented therein)* in an EagleEye field trip.	4.24	0.52
Collaboration	7. My students will have more discussion with their peers in an EagleEye field trip.	3.81	0.55
	8. My students will have better interaction with their peers in an EagleEye field trip.	3.77	0.50

Table 5.6 Subscales' reliability estimates and descriptive statistics (Part 1 of the questionnaire)

Subscale	Constructed by	Mean	SD	Cronbach's alpha
Student-centredness	Items 1, 2	4.40	0.53	.78
Motivation	Items 3, 4	4.33	0.48	.79
Scaffolding	Items 5, 6	4.24	0.54	.77
Collaboration	Items 7, 8	3.79	0.58	.70

statistics of the four subscales. The subscales' Cronbach alphas were between .70 and .79, indicating adequate reliability. We conducted three MANOVAs to assess whether there were differences respectively (i) between the three age groups, (ii) between the four teaching-experience groups, and (iii) between the three school-banding groups, on a linear combination of the four subscales. The analyses indicated there were no significant differences. In other words, no evidence showed that the teachers' age, teaching experience, or school banding influenced their responses to these eight items.

According to the results, the teachers did perceive that an EagleEye field trip works better than conventional field trips do in terms of the four subscales: Student-centredness (4.40), Motivation (4.33), Scaffolding (4.24), and Collaboration (3.79). Nevertheless, the Collaboration's mean was relatively smaller than the other three subscales'. A one-way within-subjects ANOVA further indicated that there were significant differences among these four subscales' means (Wilks' lambda = 0.71, $F(3, 95) = 25.10$, $p < 0.001$). Post hoc paired-samples t-tests evidenced that the Collaboration's mean was significantly different from the other three subscales', revealing it was the smallest. The qualitative data gathered during the training could complement the above findings:

Student-centredness. The survey results indicated the teachers perceived positively that EagleEye can offer students better student-centred learning experience than conventional field trips do. In the chat with *Teacher A*, she elaborated—*I can see the EagleEye approach shifts fieldtrip activities from a traditional, teacher-centred manner to an actively student-centred one. Students can design their own exploratory route in an EagleEye fieldtrip … they can have more control on and autonomy for exploring the real world. This is very valuable learning experience in Geography education.*

Motivation. The survey results showed the teachers realized positively that EagleEye can motivate and engage students in a greater extent than conventional field trips do. In the chat with *Teacher B*, he elaborated—*in an EagleEye fieldtrip, different student groups can frame their own exploratory route. This motivates them to think more during the fieldtrip. Besides, since the routes are framed by themselves, definitely this will increase their engagement in the activity … also, fewer number of students arrive at a single exploratory spot at the same time … this allows each student group to have more room for observing and experiencing the things on each exploratory spot.*

Scaffolding. The survey results revealed the teachers realized positively that EagleEye can better scaffold students than conventional field trips do. In the chat with *Teacher C*, he elaborated—*I can see that EagleEye can provide students with much better support than paper-based worksheets do …… compared to the worksheets used in conventional fieldtrips, the just-in-time popping-up feature of the exploratory scaffolds on a location-aware exploratory resource can let students be more aware of their current geo-locations on the site … and more effectively draw their attention to what they should probe into …… compared to static text and images on the worksheets, the multimedia presentation of the resource can offer them clearer guidance in the course of exploration.*

Collaboration. The survey results signified the teachers just moderately perceived that EagleEye can better promote collaboration among students than conventional field trips do. This might be due to the reason that in the training we provided each teacher with an iPad (not in a group basis), and hence, they might realize that we should have been suggesting an EagleEye field trip should be implemented with a "one-to-one device approach" (Looi et al. 2011). As observed in the training, although the teachers were divided into groups during the field trip (in Session 1 of the training, see Sect. 5.5.1), they did not communicate much with one another. Most of the time, they were paying attention closely to their own iPad. Thus, the teachers might perceive that their students would also perform similarly in an EagleEye field trip as they did. In the chat with *Teacher D*, she elaborated—*all youngsters are fascinated with Apple™ products …… This scenario*

Table 5.7 Descriptive statistics (Part 2 of the questionnaire)

	Item	Mean	SD
User-friendliness	The LERAT is easy to use.	4.38	0.51
	The GEP is easy to use	4.42	0.60
	The TC is easy to use	4.40	0.58

will likely happen in an EagleEye fieldtrip …… when every student is provided with an iPad, each of them may only look closely at their own device during the fieldtrip, without much discussion or interaction with their groupmates.

5.6.2 User-Friendliness of EagleEye

Table 5.7 displays the descriptive statistics of Part 2 of the questionnaire. The three items aimed at gathering the teachers' feedback on the user-friendliness of the front-ends of EagleEye (i.e. the LERAT, GEP, and TC). Similarly, we conducted three MANOVAs to assess whether there were differences respectively (i) between the three age groups, (ii) between the four teaching-experience groups, and (iii) between the three school-banding groups, on a linear combination of these three items. Again, the analyses indicated there were no significant differences, i.e. no evidence showed that the teachers' age, teaching experience, or school banding influenced their responses to these items. In addition, a one-way within-subjects ANOVA indicated that there were no significant differences among these three items' means.

The results showed that the teachers found the operation of EagleEye user-friendly, aligning with what we observed. In the training, they could complete all hands-on tasks. Just very few of them needed to approach the helpers for assistance. In the chat with *Teacher E*, she elaborated—*I am not proficient in information technology. However, I didn't have any difficulties in converting my paper-based field trip worksheets into a location-aware exploratory resource with the LERAT. I also have no difficulties in manipulating the GEP and TC. The overall operation of the system is quite user-friendly.*

When interacting with the teachers, we saw that most of them were using touch-screen smart phones (e.g. Apple™ iPhone, Android phone, etc.). Their prior knowledge in operating smart phones and playing Apps might help them get familiarized with EagleEye more easily and quickly, and thus, they found the system user-friendly.

5.7 Conclusion and Future Work

In this chapter, we have delineated our mobile learning initiative, EagleEye, an integrated GPS-supported system for supporting teachers and students respectively in facilitating and pursuing exploratory learning in Geography field trip activities. We have also discussed our study from the teacher viewpoint (with 98 Geography teachers) to evaluate our work. The results revealed that, in comparison with conventional field trips, the teachers realized that EagleEye can offer students better student-centred learning experience, enhance their learning motivation, and provide them with better scaffolds. They also found the operation of EagleEye user-friendly, without difficulties in getting familiarized with the LERAT, GEP, and TC. However, the teachers had reservation about whether the adoption of EagleEye can promote better collaboration among students in a field trip activity. As mentioned in Sect. 5.6.1, it might be due to the reason that the teachers perceived an EagleEye field trip should be implemented with a "one-to-one device approach" which hampers students' discussion and interaction during a field trip. Currently, we do not have an answer to whether an EagleEye field trip should be operated in a "one-group-one-device" or "one-to-one device" manner in order to optimize students' collaboration therein. This is an important area that we are going to investigate in the near future.

The research subjects in the present study were not a random sample. They were Geography teachers who were more or less interested in knowing more about mobile learning or/and EagleEye (that was why they enrolled and participated in the training). However, as argued by Fullan (1977, 2007), even if a teacher wants to learn more about an educational innovation, after experiencing it, he/she can still rate it negatively because of the discrepancies between his/her "expected" and "experienced" educational value of this innovation. We believe our teacher participants did provide us with their objective feedback on EagleEye.

Last but not least, we have to underline again that the teachers' feedback we received in this study was according to their experience in the EagleEye training, not their experience in implementing EagleEye in their real teaching practice. They might perceive our work differently after having real experience in facilitating their students to pursue exploratory learning with EagleEye. We will conduct further research on the aspect of teachers' implementation of EagleEye for having a more conclusive evaluation of our work.

Acknowledgments The work described in this paper was substantially supported by Mr. Eric Luk (Assistant Computer Officer) and the research assistants in the Centre for the Advancement of Information Technology in Education at The Chinese University of Hong Kong.

References

Ashbrook, D., & Starner, T. (2002). Learning significant locations and predicting user movement with GPS. In *Proceedings of the 6th International Symposium on Wearable Computers* (pp. 101–108). Retrieved December 30, 2013, from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01167224

Baker, A. C., Jensen, P. J., & Kolb, D. A. (2002). *Conversational learning: An experiential approach to knowledge creation*. Westport: Quorum Books.

Bell, P., Lewenstein, B., Shouse, A. W., & Feder, M. A. (Eds.). (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, DC: National Research Council.

Biggs, J. B., & Moore, P. J. (1993). *The process of learning* (3rd ed.). New York: Prentice Hall.

Brush, T. A., & Saye, J. W. (2002). A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. *The Journal of Interactive Online Learning, 2*(1), 1–11.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Douglass, D. (2008). *The out-of-classroom experience*. Retrieved December 30, 2013, from http://www.4faculty.org/includes/digdeeper/Outside/outside.htm

Frew, J. (1999). *Geography fieldwork projects*. London: Nelson Thordes.

Fullan, M. (1977). Research on curriculum and instruction implementation. *Review of Educational Research, 47*(1), 335–397.

Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York: Teachers College Press.

Hofstein, A., & Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. *Studies in Science Education, 28*, 87–112.

Jong, M. S. Y. (2013a). Design and implementation of EagleEye—An integrated outdoor exploratory educational system. *Research and Practices in Technology Enhanced Learning, 8*(1), 43–64.

Jong, M. S. Y. (2013b). *Student-centric eLearning: From a constructivist perspective*. Paper presented at Learning and Teaching Expo 2013, Hong Kong, China.

Jong, M. S. Y., Lee, J. H. M., Luk, E. T. H., & Lee, F. L. (2012). *The ultimate solution for outdoor exploratory learning*. Poster presented at International ICT Expo 2012. Hong Kong, China.

Knapp, D., & Barrie, E. U. (2001). Content evaluation of an environmental science field trip. *Journal of Science Education and Technology, 10*(4), 351–357.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development*. New Jersey: Prentice Hall.

Lederman, L. C. (1992). Debriefing: Toward a systematic assessment of theory and practice. *Simulation & Gaming, 23*(2), 145–160.

Lenon, B., & Cleves, P. (2001). *Fieldwork techniques and projects in geography* (2nd ed.). London: Collins.

Looi, C. K., Zhang, B., Chen, W., Seow, P., Chia, G., Norris, C., et al. (2011). 1:1 Mobile inquiry learning experience for primary science students: A study of learning effectiveness. *Journal of Computer Assisted learning, 27*(3), 269–287.

Nadelson, L. S., & Jordan, J. R. (2012). Student attitudes toward and recall of outside day: An environmental science field trip. *The Journal of Educational Research, 105*(3), 220–231.

Nordmark, S., & Milrad, M. (2012). Mobile digital storytelling for promoting creative collaborative learning. In *Proceedings of the 7th IEEE International Conference on Wireless, Mobile and Ubiquitous Technology in Education* (pp. 9–16). Takamatsu, Japan.

Ogata, H., Hui, G. L., Yin, C., Ueda, T., Oishi, Y., & Yano, Y. (2008). LOCH: Supporting mobile language learning outside classrooms. *Mobile Learning and Organisation, 2*(3), 271–282.

Pachler, N., Bachmair, B., & Cook, J. (2013). A sociocultural ecological frame for mobile learning. In Z. L. Berge & L. Y. Muilenburg (Eds.), *Handbook of mobile learning* (pp. 35–46). New York: Routledge.

Prensky, M. (2010). *Teaching digital natives: Partnering for real learning*. Thousand Oaks: Corwin.

Sharples, M. (2002). Disruptive devices: Mobile technology for conversational learning. *International Journal of Continuing Engineering Education and Life Long Learning, 12*(5/6), 504–520.

Shih, J. L., Chuang, C. W., & Hwang, G. J. (2010). An Inquiry-based mobile learning approach to enhancing social science learning effectiveness. *Educational Technology & Society, 13*(4), 50–62.

Shuler, C. (2009). *Pockets of potential: Using mobile technologies to promote children's learning*. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Vygotsky, L. (1978). *Mind and society*. Cambridge: MIT Press.

Wake, K. D., & Wasson, B. (2011). Supporting creativity in teaching and learning of history through small-group production of mobile, location-based game. In *Proceedings of the 10th World Conference on Mobile and Contextual Learning.* Beijing, China.

Wang, F., & Hannafin, M. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53*(4), 5–23.

Zurita, G., & Baloian, N. (2012). Context, patterns and geo-collaboration to support situated learning. In J. Bravo, D. López-de-Ipiña, & F. Moya (Eds.), *Ubiquitous computing and ambient intelligence* (pp. 503–511). Heidelberg: Springer.

Chapter 6
Note-Taking in Pupil's Textbook: Features and Influence Factors

Guang Chen, Chaohua Gong and Ronghuai Huang

Abstract As an important learning strategy, note-taking has been drawing researchers' attention for a long time. This research analyzes features and influence factors of note-taking in used Chinese language and Mathematics textbooks for primary schools through content analysis approach and interview approach leading to the following results. Firstly, the features of note-taking in the textbooks embody in three aspects of the forms, locations, and contents of note-taking. Secondly, the important influence factors of note-taking contain individual initiatives, teachers' lecturing speed, and the difficulty of contents. Thirdly, the note-taking contents have obvious diversities in disciplines and grade sections. Fourthly, the textbook is an important cognitive tool for pupils' study and an important mediator for family-school communication. Fifthly, teachers have to guide pupils to develop good habits of note-taking. Sixthly, the design of note-taking function for e-textbook should take the features of note-taking in printed textbooks into consideration with the enrichment of note-taking forms and providing different note-taking tools, etc.

Keywords Note-taking · Annotation · Textbook · e-textbook

G. Chen · R. Huang
School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing, China

C. Gong (✉)
School of Computer and Information Science, Southwest University, Chongqing 400715, China
e-mail: gongchaohua@gmail.com

G. Chen · C. Gong · R. Huang
Collaborative and Innovative Center for Educational Technology (CICET), Beijing Normal University, Beijing, China

G. Chen · R. Huang
Beijing Key Laboratory of Education Technology, Beijing Normal University, Beijing, China

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_6

6.1 Introduction

For a long time, taking notes has been widely used in the instruction as an important learning method, and both teachers and students have deeply understood the benefit of taking notes to the learning effects of students during the long-term process of learning and teaching. As the important significance of note-taking was found in the field of education, many educational psychologists have carried out systematic studies on note-taking and have many research achievements in this field (Yang 1988). The achievements of former researchers mainly focus on the following areas.

Studies on the function of note-taking: There are storage function hypothesis and encoding function hypothesis. Storage function hypothesis regards the note-taking as the carrier of knowledge, which facilitates the learner's memorization and retrieval. Encoding function hypothesis advocates that the process of taking notes can cause positive activities in learning, which is conducive to organizing memory and forming migration (Kiewra 1989; Kiewra et al. 1991; Rickards and Friedman 1978).

Studies on the theory of taking notes: There are quantitative theory and qualitative theory. The qualitative theory focuses on "which contents are in the learners' concern and taken in their notes," while the quantitative theory focuses on "why the statements taken down can be well memorized." Both give explanation to the occurrence of the effects of taking notes (Peper and Mayer 1986).

Studies on the technical issues of taking notes: In accordance with the defects in the note-taking method of taking verbatim notes of the teacher's words in the traditional way, (Kiewra 1985; Kiewra et al. 1991) proposes two theoretical hypotheses for determining the merits of note-taking methods, which are the degree of integrity and that of internal connections. He also puts forward two methods of taking notes, which are linear technique and matrix technique.

Studies on the influence factors of note-taking: The researchers have found out that the quality and quantity of taking notes are influenced by the lecturing speed, the knowledge background of the subjects, and the length of lectures, etc. (Chen and Chiu 2011).

Studies on the note-taking in textbooks: The note-takings in the textbooks mainly are annotation notes containing two forms. One is mark such as underlines and circles and the other is text annotation (i.e., annotations between the lines, a separate note excerpt) (Marshall 1998). Besides, there are studies on notes sharing, the analysis of the functional features of e-note system (Hoff et al. 2009) and so on.

The above studies have achievements in the aspects of the function of note-taking, the ways and techniques of taking notes.

However, so far, most of these studies were carried out in developed countries such as USA. There have been very few studies on the issues of taking notes in developing countries such as China. Researchers find out that the pedagogical content knowledge in these two countries differs markedly, which has a deep impact on teaching practice (An et al. 2004). Based on the above background, this study

Table 6.1 The correspondent relations between the forms and functions of different annotation note-takings

Forms	Functions
Underline or highlight chapter catalog and headings; personalized note-taking symbols (such as advanced asterisk, strikethrough)	To remind him/her of paying attention in the future
Short highlights, circled words, drawing between the lines of the texts), special symbols at paragraph boundary (use asterisk)	To mark location and facilitate memorization
Notes on the sidebar, charts, or near the equations	To solve problems
Notes on the sidebar, in the line spacing of the texts, words, and phrases	To interpret
Long highlights or underlines	To use complex descriptive tracking process
Notes, drawing, graffiti, and other marks that are unrelated to the content itself	Reflection along with the process of reading materials

focuses on the features of pupils' note-taking in their textbooks and the influence factors of taking notes, with the emphasis on the analysis of the used Chinese language and Mathematics textbooks for primary schools published by People's Education Press, China and Beijing Normal University Press, China. It carries out content analysis approach to analyze the note-taking left in the textbooks by the pupils and acquires the important influence factors of taking notes through interview approach. The results of the study aim to provide references to the design of note-taking function for e-textbooks. Therefore, this study proposes the following research questions:

(1) Which features do the notes taken in the printed textbooks by pupils have?
(2) What are the factors that influence the note-taking of pupils?

6.2 Literature Review

6.2.1 Note-Takings in Printed Textbooks

Marshall (1998) collects 410 used textbooks from the campus rental bookstore. After thorough investigation, he summarizes the forms of the annotation notes in these textbooks and proposes corresponding functions (see Table 6.1). In addition, from the attributes of the note-taking, he also finds out that the notes in these textbooks contain formal classroom note-takings, private tips during self-reading, explicit key knowledge, and implicit learning experience, and so on.

In addition, Wang (2009) summarize that there are two kinds of notes in the existing textbooks. One is the note added by the editor in order to facilitate students' learning, which can be divided into three types: question notes, association notes, and reflection notes. The other one is the notes taken autonomously by teachers and students in their reading process, which includes two forms: (a) annotations such as underlining and circling; (b) wordings are usually being written in-between lines or written as sidebar notes.

Although Marshall's study involves the study on the note-taking in the textbook, he only takes college students as the subjects without concerning about the forms of note-takings of K-12 school students. So, the studies on how K-12 school students take notes in the textbook, what features their note-takings have and what are the influence factors of pupils' taking notes have not drawn much attention of the researchers.

6.2.2 *Computer-Based Annotation Systems*

Hoff et al. (2009) analyze the existing computer-based annotation systems from four dimensions of media-supported formats, annotation support, annotation management, and user interaction and sort out the common attributes of them, as shown in Table 6.2. However, there are several shortcomings in these annotation systems, including "limitation to single document formats, no capturing of relations to reflect lateral reading, loss of context information an annotation was created in, and limited sharing capabilities among collaborative users."

Table 6.2 Schematic overview of the characteristic properties of computer-based annotation systems (Hoff et al. 2009)

Supported formats	Document	Web documents (HTML), office documents (MS Office, Open Office, etc.), multimedia (graphics, video, audio)
	Annotation	Text, graphics, audio, video, link
Annotation	Placement	In-place, on-place, off-place
	Locality	Whole document, predefined elements, arbitrary elements, multiple elements/documents
	Semantic	Mark-up elements (e.g., question, reminder, etc.)
Management, interaction	Shareable	Private, group, public
	Reply possibility	Discussion threads, "annotating annotations"
	Notification	Notify author about new annotation, reply notification (e.g., through email)
	Searchable	Annotations, metadata, natural language understanding system
Implementation	Repository	local, global, replicated (client–server, peer-to-peer)
	Technology	Reference points for annotations (e.g., XPointer, RDF, hash bashed), versioning

Cabanac et al. (2007) also point out that due to the inappropriate setting of the metadata, the electronic annotation systems may not correctly express the intention of the annotation takers.

Besides, McFall (2005) attempts to carry out metaphor design of e-textbook note-taking function in the way of that in printed textbooks and adds the functions of deleting and modifying note-takings by using the function of management of electronic note-taking system, but this system does not bring any advantages to textbook reading from the perspective of application result.

6.3 Research Methodology and Design

This research aims to analyze the features and influence factors of note-takings in Chinese and Mathematics textbooks in K-12 schools in China. It puts forward the analytical dimensions of note-taking features by literature study, studies on the features of the existing note-takings, and expert consulting process so as to analyze and compare the features of note-takings by calculating the number of pages containing note-takings in the textbooks. In addition, it analyzes the main influence factors of note-takings through interviewing K-12 school students for their understandings of the note-taking.

6.3.1 The Selection of Textbooks

The objects of this research are the used Chinese language (as shown in Fig. 6.1) and Mathematics textbooks for primary schools published by People's Education Press, China and Beijing Normal University Press, China, that are widely selected currently. The range of selected grades is from Grade 1 to Grade 6 with a total of 50 textbooks.

6.3.2 Research Tools

Referring to the research achievements of Marshall (1998) and Hoff et al. (2009), this research categorizes based on the specific circumstances of the behavioral features of note-takings in the textbooks for primary schools. Ou and Yang (1998) thinks that the analytical unit of textbooks can be chapters, sections, units, lessons, paragraphs, words, sentences, characters, pages, and any other units. It is not suitable for this research to be over specific. In order to accurately master the materials for objective analysis, the research uses "page" as the unit of analysis. The interview outlines of topics of this research are made up mainly referring to the studies of Liu (2003) and so on.

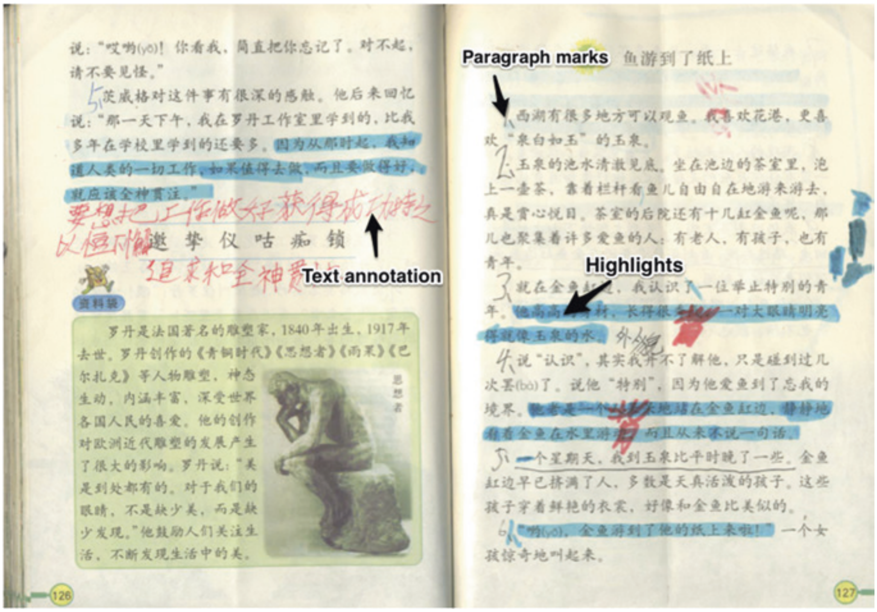

Fig. 6.1 One of the used Chinese language textbooks with 3 kinds of annotations (*text annotation*, *paragraph marks*, and *highlights*)

6.3.3 Reliability and Validity

There are two raters to take reliability tests and they are graduate students in the major of "Curriculum and Teaching Theory" and go to K-12 schools to observe classroom teaching for investigation so as to get familiar with the instruction of Chinese language and Mathematics.

In the reliability test process, the researcher first selects 50 pages, distributes the classification tables to the two raters, then describes the principles and methods of classification, asks them to classify the selected pages based on the contents accordingly, and finally uses the equation to calculate the reliability based on the results of classification. The reliability analysis equations used in this research are as follows (Kelly et al. 2008):

(1) Degree of mutual agreement

$P = 2M/(N_1 + N_2)$

M number of totally agreements;
N_1 number of expected agreements of the first scorer;
N_2 number of expected agreements of the second scorer

(2) Reliability = T * (Average degree of mutual agreement)/(1 + (T − 1) * (Average degree of mutual agreement))

T number of raters

According to the above equations, the reliability of this research is 0.88, which meets the requirements of content analysis. In the perspective of validity, the researcher has repeated discussions with three Chinese language teachers in K-12 schools and an expert of curriculum and instructional theories, and ensures to cover all important contents through constant modifications so as to satisfy the requirements of this research.

6.4 Results

6.4.1 Features of Note-Takings on the Textbooks for Primary Schools

Based on statistics of the note-taking presented in the textbooks, we summarize the results of the collected samples in three dimensions of the note-taking forms, locations, and contents, as shown in Table 6.3. The features of these note-taking reflect that Chinese language emphasizes on helping students to understand the text content and to improve their Chinese language learning capability by circling, drawing, and commenting the characters, words, sentences, and paragraphs.

As shown in Table 6.4, Mathematics focuses on help students to understand the knowledge of numbers and algebra, statistics and probability, graph transformations and space, etc., by mathematical language symbols and annotations.

Table 6.3 Results of the features of note-takings on the Chinese language textbooks

Dimensions	Chinese language
Note-taking forms	Paragraph marks (numbers before the paragraph, double vertical bars after the paragraph)
	Underlines (straight lines, wavy lines, vertical bars between sentences)
	Highlights
	Annotation marks (geometric images, dots, brackets, numbers)
	Pinyin marks
	Graphic marks (circles, dots)
	Text annotation (copying homework, notes for the text)
	Color fill
	Proofreading marks
Note-taking locations	Eyebrow notes (at the margin of the headlines)
	Margin notes (next to the new characters and words in the text, next to the vocabulary table)
	Interlinear notes (labels before the paragraph, add annotations, or notes to the text contents)
	Notes beside the text (write down the main contents of the paragraph on both sides of the margins, figures of speech, sentence paraphrase)
	Notes at the margin below the text (after the paragraph)
	Notes at the classic words or sentences
	Fly page (homework, personal information)

(continued)

Table 6.3 (continued)

Dimensions	Chinese language
Note-taking contents	Learn new characters and words: make up phrases, phonetic records, phonetic new characters
	Write characters: square lattice, practice strokes, stroke orders, copying the phonetic alphabets (Chinese characters)
	Grammar: punctuation, rhetoric, sentence analysis (vertical bars)
	Understanding of the texts: the main idea, explanation of characters and words, notes of the idea of paragraphs
	Homework: note the homework, answers to homework
	Teacher's comment on assignment
	Signatures of parents, teachers, and students
	Content integrity: the incompleteness of existing notes, misspelling
Others	Sticker notes

Although the note-takings are of various forms and vary to different subjects, the purposes and functions of note-takings are more consistent, which record the combined the reading, thinking, and memorizing process of learners to enhance their understandings of teaching materials for review and knowledge transference so as to promote effective learning.

We further sort out the corresponding relations between the note-taking forms of Chinese language and pupils' grade sections and the statistical results are shown in Fig. 6.2. The figure shows that the note-taking forms are closely related to the grades of the pupils. The item of color fill only appears in Grade 1 and Grade 2. The items of highlight and graphic marks do not appear in the textbooks until Grade 3. The items of using sticker notes and proofreading marks for sentences rarely appear in the textbooks, which can be found only in Grade 4. The appearance frequency of Pinyin marks decreases as the grades increase.

We sort out the forms and contents of note-taking in the Mathematics textbook and the statistical results are shown in Fig. 6.3. The frequency of the note-takings in the Mathematics textbooks reflects the subject feature of Mathematics and the note-taking contents are closely related to the solving process of mathematical problems. Furthermore, Mathematics textbooks are also the important media for teacher–student interaction, as well as teachers' comment on students' assignment, which is particularly significant in the lower grade sections.

6.4.2 The Influence Factors of Pupils' Taking Notes

After the interview of 18 pupils in Grade 4 to Grade 6, we find out that they generally agree that the note-takings provide support to their review in the future, so they will be initiative to take notes. Teachers often ask students to take notes and sort out the notes, but rarely examine the notes. Students generally think that the speed of teacher's lecturing is an important factor that influences their note-takings. If the teacher leaves them very little time, they will have incomplete

Table 6.4 Results of the features of note-takings on the mathematics textbooks

Dimensions	Mathematics
Note-taking forms	Mathematical symbols (<, <, ≠, =, etc.)
	Circle marks
	Drawing with a variety of learning tools (triangle, rectangle, square, circle, etc.)
	Mathematical text notes (notes for equation, equation classification)
	Painting and drawing lines: painting, filling in the colors, matching
Note-taking locations	Interlinear notes (among the application problems, among the mathematical problems)
	Notes beside the text (between the margins in the exercises)
	Fly page (memo, homework, personal information)
Note-taking contents	Calculation steps (addition, subtraction, multiplication, division, multi-step calculation)
	Annotations of the numbers (algebra and geometry calculative results)
	Answers to the exercise (the answers to application questions and blank-filling questions)
	Measurement results of the learning tools (rulers, triangle rulers, compasses, etc.)
	Circle the exercise
	Teacher's comment on assignment
	Assignments
	Signatures of parents and teachers
Others	Sticker notes

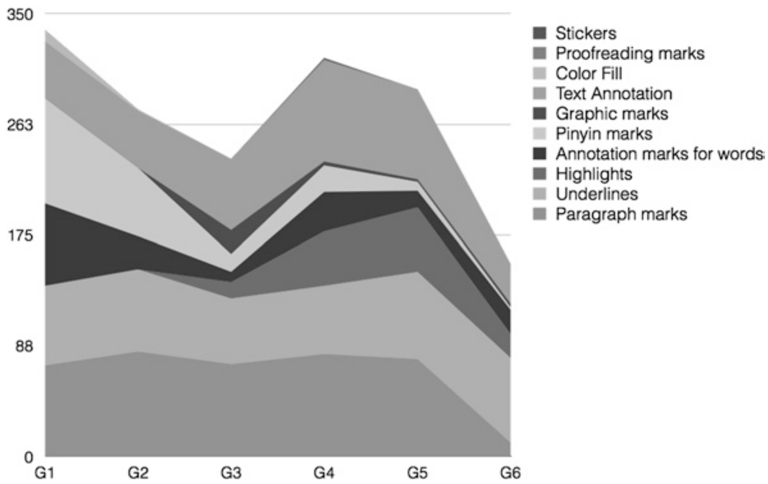

Fig. 6.2 Results of the forms of note-takings in Chinese language textbooks for different grades (Unit: Page)

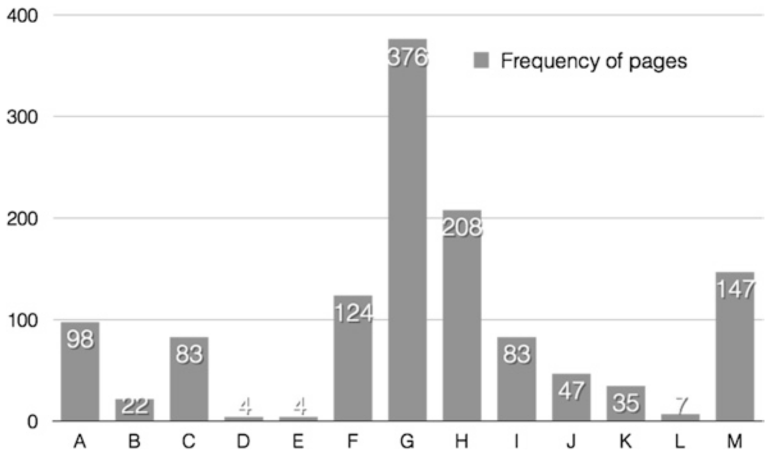

Fig. 6.3 The statistical results of the contents and forms of note-taking in Mathematics textbooks (Unit: Page). *Notes A* = Calculation steps; *B* = Stickers; *C* = Assignments; *D* = Note-taking text errors; *E* = Measurement results of the learning tools; *F* = Text interpretation; *G* = Fill in the blanks with numbers; *H* = Answers to the application problems; *I* = Circle marks; *J* = Matching; *K* = Fill in the blank with symbols; *L* = Autonomous painting; *M* = Teacher's comment on assignment

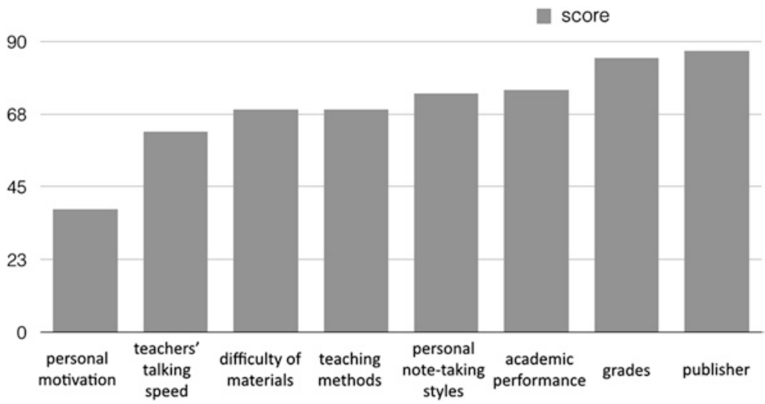

Fig. 6.4 The ordering figure of the influence factors of pupils' taking notes

note-takings. In terms of the difficulty of the text contents, students all agree that their note-takings are usually more detailed with more information when encountering with learning contents of low relevance to their existing knowledge. The subjects list out the orders of the influence factors of note-takings according to the degree of importance as shown in Fig. 6.4. The lower the score is, the higher the degree of its importance is.

Therefore, among the eight influence factors of pupils' note-takings, four factors are related to the learners themselves, which are personal initiatives, personal note-taking styles, learning scores, and the grades of pupils, respectively; two are related to teachers, which are the speed of teachers' lecturing and teaching methods; and two are related to the textbooks, which are the difficulty and edition of the contents. The personal initiatives, the speed of teachers' lecturing, and the difficulty of the text contents are the most important influence factors. Most students do not think that the edition/publisher of textbooks has any influence on the note-takings.

6.5 Discussion

6.5.1 Features of Note-Taking Are Subjects Related

By analyzing the behavioral features of the collected pupils' note-takings, this research finds out that the focuses of note-taking contents of different subjects have significant differences, so do the note-taking contents of different knowledge points within the subject.

The note-takings of Chinese language are mainly about learning new characters and words and understanding contents with a large number of text notes and the massive use of various symbols such as circles and underlines. Pupils annotate the texts of different literary genres from different aspects. The annotations of narratives focus on the environment, plot, characters, theme, and other aspects, those of proses focus on the thoughts and feelings, the moods, artistic conception and language, etc., while those of the poetry reflect their feelings of the poems, the images in the poems, subtle meanings, poetic languages, and other esthetic and appreciation capabilities.

The note-takings in the arithmetic process of Mathematics show the learners' utilization of addition, subtraction, multiplication, and division, which truly reflect their problem-solving process. In terms of the application problems, the note-takings of circling the important information of the problems reflect their understanding process of the problems. Therefore, the process of taking notes and the contents of note-takings are important cognitive strategies for pupils to grasp key information.

6.5.2 Note-Takings of Chinese Language Are Grade Section Related

The Chinese curriculum standards divides Chinese language for primary schools into three grade sections, which are Grade section one (Grades 1–Grade 2), Grade section two (Grade 3–Grade 4), Grade section three (Grade 5–Grade 6). The differences of note-taking features in Chinese language are obvious in terms of grade sections and the results of sorting out are as follows.

In Grade section one, the note-taking form is mainly personalized symbols, such as circling characters and words, underlining words and sentences, and writing down serial numbers of paragraphs, as well as plenty of various irregular special graphics and symbols. Note-taking contents mainly make phonetic to the new characters and make up phrases with them. The most important point reflected by the note-taking is that learning new characters and words is the key learning task in Grade section one. The teachers need to motivate pupils in learning and writing characters. Besides, the note-taking contents have strong correlation with exercise and assignment, as the note-taking in some pages are mainly the pupils' completing their exercise or homework. So the features of pupils' note-taking behaviors at this stage are not steady and their note-taking contents only reflect their concerns about the key knowledge points of the text without the establishment of reinforced practice. The note-taking contents also reflect their completing the exercise or homework. So the proportion of the total number of exercise is the highest in the three grade sections.

In Grade section two, the style of note-takings in different textbooks show some common features, such as highlight key characters, words, and sentences, and writing comments to the contents from the characters and words to sentences and paragraphs. The content of notes reflects their understandings of the text contents, for example, note-takings show the pupils' learning to summarize the main idea of paragraphs and conclude the main idea of the whole text, and also penetrates their understanding and application of some grammar points. They begin to extract some good words and good sentences, which indicates the gradual transition from the accumulation of characters and words in the first-grade section to that of phrases and sentences.

Compared with Grade section two, the note-taking in the textbooks in Grade section three shows less accumulation of single characters and words but more application of characters and words and the understanding of text contents and structures, especially some note-takings reflect that the pupils can summarize the main contents of different articles and the development of story plots by outline note-takings. The note-takings at this stage reflect in the whole that the note-taking habits of the learners are preliminarily formed. The note-taking contents and forms better reflect their mastery of the knowledge points. The note-taking locations are closer to the associated knowledge points. So, the note-taking behavior is transforming into an effective learning strategy.

6.5.3 Textbook as an Cognitive Tool and an Intermediary Tool

We find out that teachers, pupils, and parents use textbooks as the intermediary for family-school communications, exchange activities among pupils, and between teachers and pupils from the analysis of note-taking. The lower the grade is, the more frequent the family-school communication is. Conclude from further analysis, teachers' activities are mainly reflected in correcting pupils' homework, leaving comments to parents, signing below the texts that require pupils to recite and so on; while parents' activities are mainly reflected in the records of instructing their children to complete homework and sign after checking the homework; and the interaction among students mainly refers to the records of reciting texts in peers.

Therefore, in the current instruction activities in schools, the textbook is an important intermediary tool for family-school communications. Parents acquire the information from the teachers by examining their children's textbooks and record the information of supervising their children to complete homework in the textbooks, which allows the teachers to have a full understanding when examining the textbooks. So the analysis of the features of note-taking reflects that the textbook does not only carry the function of an important cognitive tool for pupils' study, but also an important intermediary for family-school communications.

6.5.4 Students' Habit of Taking Notes Depends on Teachers

As an important learning strategy, note-taking has been attached great importance to both teachers and students. The influence factors of taking notes are diverse with students, teachers, and contents as the important influence factors of taking notes, among which personal initiative plays a key role.

Pupils are in the critical period of forming good study habits and the amount of note-takings in class is the largest out of their note-takings. Thus, teachers need to guide them to form the good habit of taking notes, for instance, training their capability of taking notes in the texts of different literary genres, standardizing the note-taking by specifying unified symbols, learning the techniques of taking notes from one another by note-taking sharing activities, instruct them to take notes efficiently by providing some note-taking tools, etc. The teachers ought to help them to transform from passive note-taking to active note-taking so that the process of taking notes and note-taking contents can effectively promote their learning.

6.6 Implications for e-textbooks

How to change the ways of learning and teaching in the information age is the key issue that the researchers of K-12 education should concern about at present. As an important component of the change in the ways of learning and teaching, printed textbooks are clearly insufficient in response to the process of education reform (Huang et al. 2010), so the e-textbooks are put on the agenda as a new form of textbooks (Chen et al. 2012; Gong et al. 2011). In general, e-textbooks refer to a type of e-books with special functions that meet specific content standards, which are lack of systematic and detailed studies. Although some electronic textbooks or equipment have appeared in the classroom, they have not fully prepared for classroom instruction (Weisberg 2011).

According to the investigation conducted earlier by some scholars (Morton et al. 2007), the results show that teachers and students consider the functional features of e-textbooks as the key to the process of the construction of e-textbooks in the future and the note-taking based on the textbooks is considered as an important tool with which the pupils complete their self-study and knowledge construction, so the design of note-taking function for e-textbooks is very important. At present, some existing e-textbooks have poor user experience, which simply cannot serve as an alternative other than printed textbooks. The main reason is that the design of note-taking function does not meet the needs of learning and teaching, resulting in the less natural appearance than in the printed textbooks.

The analysis results of note-taking of printed textbooks have significant meaning of guidance in practice for the design of note-taking function for e-textbooks, according to which, we believe that the design of note-taking function for e-textbooks should consider the following aspects:

(1) Use recording software to enrich note-taking forms: In the classroom learning process with printed textbooks, the speed of teachers' lecturing influences students' taking notes. So, in the design of note-taking function for e-textbooks, in addition to providing the functions of highlight, underline, and comments, the functions of audio recording and video recording with split screens should also be taken into consideration to improve the efficiency of note-taking, so that students can concentrate on classroom learning.
(2) Design the sharing function of note-taking: Traditional notes are usually taken in the textbooks, which become the cognitive tool of personal review. The teachers barely have time to examine the situations of the note-takings of the students and the students seldom borrow note-takings from each other, which leads to the errors of common sense in some parts of the note-taking contents and the difficulty in sharing the techniques and experience of taking notes. Thus, e-textbooks can consider designing the function of note-taking sharing to help with learning, borrowing, and commenting on the note-takings between teachers and students or among students themselves so as to promote their overall understanding of the knowledge.
(3) Note-taking synchronization and backup: Traditional note-takings are dependent on the textbooks, so in case the textbooks get lost, so will the note-takings. As there is a wide range of the types of electronic reading devices, in order to support free reading on different devices, we must design the functions of synchronization and backup in the design of note-takings for e-textbooks, which can realize browsing on different platforms and protect note-takings in real-time synchronization and secure storage.
(4) Note-taking management: Traditional note-takings are usually stored together with the textbooks after finish learning them, so it is difficult to track the students' note-takings. The note-taking functions of e-textbooks can consider updating the note-takings, automatically detecting the mistaken characters and generating note-taking data of the learners to help them effectively manage their note-takings so as to play the important role of note-takings.

A variety of supporting tools for taking notes: Due to the significant differences of note-takings in different subjects and grade sections, the note-taking functions of the e-textbooks have to provide a variety of note-taking techniques and tools to help learners to complete study the courses of different subjects in different grades.

6.7 Conclusions

This research analyzes features and influence factors of note-taking in Chinese language and Mathematics textbooks for primary schools. The results could answer the questions we proposed at the beginning of this chapter.

Which features do the notes taken in the printed textbooks by pupils have? We found that the features of note-taking in the textbooks embody in three aspects of the forms, locations, and contents of note-taking; the note-taking contents have obvious diversities in the subjects and grade sections and the textbook is an important cognitive tool for pupils' study and an important intermediary tool for family-school communication.

What are the factors that influence the note-taking of pupils? We found that the important note-taking influence factors contain individual initiatives, teachers' lecturing speed, and the level of contents. Therefore, teachers should guide pupils to develop good habits of note-taking.

The design of note-taking function for e-textbook should take the features of note-taking in printed textbooks into consideration with the enrichment of note-taking forms and providing different note-taking tools, etc.

Of course, as e-textbook is still new, we need to explore the specialty of its functions and optimize the design of note-taking functions based on borrowing from the features of note-takings in printed textbooks, so as to make the utilization of e-textbooks obey the laws of learning and teaching, and also promote the learners to carry out spontaneous, collaborative, and inquiry learning to meet the requirements of learning in the future.

Acknowledgments This research work is supported by National Social Science Foundation of China 2011 project "Curriculum: e-Textbook for middle school" (Fund No. CCA110107).

References

An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school, mathematics teachers in China and the U.S. *Journal of Mathematics Teacher Education, 7*(2), 145–172. doi:10.1023/B:JMTE.0000021943.35739.1c.

Cabanac, G., Chevalier, M., Chrisment, C., & Julien, C. (2007). Collective annotation: perspectives for information retrieval improvement, RIAO 2007. In *Presented at the in Proceedings of the 8th Conference on Large-Scale Semantic Access to Content (Text, Image, Video and Sound), Pittsburgh, PA*. Retrieved from http://riao.free.fr/papers/99.pdf.

Chen, C.-H., & Chiu, C.-H. (2011). Integrating blogs into elementary classrooms: shared note taking and questioning review (Vol. 5, pp. 1627–1631). *Presented at the Proceedings of 2011, 15th Global Chinese Conference on Computers in Education* (GCCCE2011).

Chen, G., Gong, C., & Huang, R. (2012). E-textbook: definition, functions and key technical issues. *Open Education Research, 18*(2), 28–32. (in Chinese).

Gong, C., Chen, G., Huang, R., & Zhang, J. (2011). The development process of e-textbook for k-12 schools in South Korea and its inspiration to China. In *Presented at the 2011 International Conference on Electrical and Control Engineering, Yichang, China* (in Chinese).

Hoff, C., Wehling, U., & Rothkugel, S. (2009). From paper-and-pen annotations to artefact-based mobile learning. *Journal of Computer Assisted learning, 25*(3), 219–237. doi:10.1111/j.1365-2729.2008.00297.x.

Huang, R., Chen, G., Zhang, J., & Wang, Y. (2010). research on informationization learning mode and its digital resource form. *Modern Distance Education Research*, *6*, 68–73. doi:CNKI:SUN:XDYC.0.2010-06-018 (in Chinese).

Kelly, A. E., Lesh, R. A., Baek, J. Y., et al. (2008). *Handbook of design research methods in education: innovations in science, technology, engineering, and mathematics learning and teaching*. New York, London: Routledge.

Kiewra, K. A. (1985). Investigating note taking and review: A depth of processing alternative. *Educational Psychologist, 20*(1), 23–32. doi:10.1207/s15326985ep2001_4.

Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review, 1*(2), 147–172. doi:10.1007/BF01326640.

Kiewra, K. A., Dubois, N. F., Christian, D., Mcshane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of Educational Psychology, 83*(2), 240–245.

Liu, Y. (2003). Cognitive pscyholinguistic approach to the function of note-taking. *Modern Foreign Language, 26*(2), 193–199. (in Chinese).

Marshall, C. C. (1998). Toward an ecology of hypertext annotation. In *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia: Links, Objects, Time and Space—structure in Hypermedia Systems: Links, Objects, Time and Space—structure in Hypermedia Systems* (pp. 40–49). New York, NY, USA: ACM. doi:10.1145/276627.276632.

McFall, R. (2005). Electronic textbooks that transform how textbooks are used. *The Electronic Library, 23*(1), 72–81. doi:10.1108/02640470510582754.

Morton, D. A., Foreman, K. B., Goede, P. A., Bezzant, J. L., & Albertine, K. H. (2007). TK3 eBook software to author, distribute, and use electronic course content for medical education. *Advances in Physiology Education, 31*(1), 55–61. doi:10.1152/advan.00036.2006.

Ou, Y., & Yang, H. (1998). *Transformation of curriculum in new century—from the perspective of Mainland China and Taiwan*. Taipei: Wu-Nan Book Inc. (in Chinese).

Peper, R. J., & Mayer, R. E. (1986). Generative effects of note-taking during science lectures. *Journal of Educational Psychology, 78*(1), 34–38. doi:10.1037/0022-0663.78.1.34.

Rickards, J. P., & Friedman, F. (1978). The encoding versus the external storage hypothesis in note taking. *Contemporary Educational Psychology, 3*(2), 136–143. doi:10.1016/0361-476X(78)90020-6.

Wang, M. (2009). Functions of Annotation Notes in Reading. *Reference for Elementary Education*, (4), 24 (in Chinese).

Weisberg, M. (2011). Student attitudes and behaviors towards digital textbooks. *Publishing Research Quarterly, 27*(2), 188–196.

Yang, Z. (1988). Review on cognitive researches of note-taking in classroom. *Advances in Psychological Science, 2*, 36–39. (in Chinese).

Chapter 7
Effects of Clicker and Peer Discussion on Learning Performance in a Secondary Biology Course

Yongbin Hu and Ronghuai Huang

Abstract From the reports of previous studies, clickers have been recognized as an effective strategy that has a positive impact on learning performance, comparing to lecture. However, it is uncertain that whether it is the clickers themselves or the more general active learning strategies that influences learning performance. The purpose of this study was to compare the effects of the two different strategies (clicker vs. peer discussion) that support active learning on learning performance of secondary students. It focused on eighty-eight secondary school female students from four classes who participated in a biology course taught by the same teacher with the same learning content, schedule, presentation slides, and questions. This research randomly selected two classes as the experimental group and the other two classes as the control group. The students in the experimental group were taught with clickers along with lecture, while those in the control group were taught with peer discussion along with lecture. Both groups completed the same tasks. Four lessons were selected in a biology course with the duration of each class about 45 min. The experiment lasted for 4 weeks. Four data collection instruments were used to compare the students' learning achievements, learning interests, learning attitudes, and cognitive levels. The results showed that (a) there was no significant difference in students' learning achievements; (b) the experimental group showed higher learning interests than the control group; meanwhile, there was no significant difference in terms of the learning attitudes in between the two groups; and (c) the experimental group showed lower cognitive level than the control group.

Keywords Active learning · Active pedagogical approaches · Clickers · Peer discussion · Secondary biology class

Y. Hu · R. Huang (✉)
School of Educational Technology, Faculty of Education,
Beijing Normal University, Beijing, China
e-mail: huangrh@bnu.edu.cn

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_7

7.1 Introduction

Over the past decades, information and communications technology (ICT) has become an important aspect of teaching and learning at all levels of education—from primary schools to universities (Tai 2011). Large-size screen, interactive whiteboard, mobile devices, as well as wireless networks are used to enrich the pedagogical approaches, offering educational institutions a unique opportunity to increase student motivation and enhance learning outcomes (Roblyer and Wiencke 2003). Because of the significant need to facilitate effective learning, it is not surprising that different teaching approaches enhanced by technology have been developed within the past 30 years.

In this study, we examine clickers, which have been adopted by many educational institutions or schools for the purpose of improving learning performance. Clickers, also called audience response systems, voting machines, wireless keypad response systems, or classroom communication systems, are interactive remote response devices that transmit and record student responses to questions providing immediate feedback to both the students and the teacher about the learning process (Homme et al. 2004). Clickers help instructors actively engage students during the entire class period, gauge students' level of understanding of the material being presented, and provide prompt feedback to students' questions for instructors. Beatty (2005) explained why clickers help students actively engage in the learning process and how this engagement helps students develop a more solid, integrated, and useful understanding of concepts and their interrelationships and applicability. A concerted focus on understanding rather than recall, on reasoning rather than answers, bolsters the effects.

In the past, much research was conducted to explore the effects of clickers on students' learning performance. Chen et al. (2010) found that the use of clickers is a popular means by which to develop an interactive environment and engage students. Clickers have been shown to increase student cognition, and metacognition may be improved (Mayer et al. 2009), which in turn will contribute to learning outcomes (Brady et al. 2013). Extensive previous research has revealed that the use of clickers has positive effects on student's learning outcomes (Beatty et al. 2006; Caldwell 2007; Duncan 2006). However, the majority of research on learning performance has only compared the use of clickers to traditional lecture methods, which is characterized as passive learning (Addison et al. 2009; Stowell and Nelson 2007; Preszler et al. 2007; Lee and Bainum 2006). Although learning performance is often higher when clickers are used, the big question is that whether clickers or the other active learning strategies are the real cause. For this reason, this study introduces peer discussion as another active learning strategy to verify the question mentioned before.

Therefore, the main purpose of this study was to compare the effects of clicker and peer discussion on the learning performance of students in secondary biology course. It addressed the following three research questions:

Do the students who learn with clickers show greater learning achievements than those who learn with peer discussion?

Do the students who learn with clickers show higher learning interests or attitudes toward the biology course than those who learn with peer discussion?

Do the students who learn with clickers show higher cognitive levels than those who learn with peer discussion?

As outlined in this paper, an experiment was conducted to evaluate the student's learning achievements, interests, attitudes, and cognitive levels.

7.2 Literature Review

7.2.1 Active Learning

Active learning, often presented or perceived as a radical change from traditional instruction, refers to several models of instruction that focus the responsibility of learning on learners. Active learning has attracted strong advocates among faculty and school teachers looking for alternatives to traditional teaching methods, while skeptical faculty regards active learning as another one in a long line of educational fads. Bonwell and Eison (1991) have contributed heavily to its development and to the acceptance of active learning as a viable approach. They defined active learning as any instructional method that engages students in the learning process. In short, active learning requires students to do meaningful learning activities and think about what they are doing (Bonwell and Eison 1991).

Using active learning does not mean abandoning the traditional activities in spite of the fact that it does take lecture time. In some cases, traditional activities such as preview before class, lecturing and homework could be introduced into the classroom as well. The core elements of active learning are students' activity and their engagement in the learning process. Active learning is often contrasted to passive learning, which is the traditional lecture where students passively receive information from the instructor (Prince 2004).

Proponents of active learning describe a process in which students engage in "doing things and thinking about what they are doing" in the classroom (Bonwell and Eison 1991). Students and their learning needs are the center of active learning. There are a lot of instructional strategies that can be employed to actively engage students in the learning process, including clicker feedback, peer discussion, problem solving, case studies, role plays, journal writing, and structured learning groups. The benefits of using such strategies include improved critical thinking skills, increased retention and transfer of new information, increased motivation, and improved interpersonal skills (Bonwell and Eison 1991).

In order to have a positive effect on students, the instructor must apply the principles of active learning to the practical setting of the classroom. Auster and Wylie (2006) suggest that four dimensions are necessary to create a systematic approach to promote active learning in the classroom: context setting, class preparation, class delivery, and continuous improvement. Context setting refers to creating an open and relaxed atmosphere for learning in the classroom. Class preparation involves thought, planning, and creativity before the class session.

Class delivery refers to the implementation of the planned lesson in the classroom. Continuous improvement entails seeking and using feedback concerning the teaching approach. Meyers and Jones (1993) have maintained that the basic elements of active learning are talking, listening, reading, writing, and reflecting.

7.2.2 Peer Discussion

Peer discussion, another active learning strategy, has been studied extensively in the K-12 classroom environment. Much of this work has focused on how students exchange ideas, disagree with one another, or support their ideas with reasons (Knight et al. 2013). Peer discussion is a process whereby a group of people express, clarify, and pool their knowledge, experiences, opinions, and feelings. Through peer discussion, certain areas of one's particular viewpoint are compared and contrasted with others.

The peer discussion class is intended to be a free give-and-take between teachers and students and among students on the current topic of concern in the course. It is characterized by probing questions from the teacher designed to elicit student interpretations, opinions, and questions. Discussion requires the learners to think critically on the subject matter and use logic to evaluate their and others' positions. As learners are expected to discuss material constructively and intelligently, discussion is a good follow-up activity given the unit has been sufficiently covered already (McKeachie and Svinicki 2006). Abdu-Raheem (2010) recommended that as a matter of urgency, teachers of social studies should explore and use discussion method of teaching to promote knowledge to secondary school students. She explained further that if discussion method is properly managed, it will go a long way to improve students' achievements in social studies.

In terms of the effects of peer discussion, it is generally assumed that active engagement of students during discussion with peers, some of whom know the correct answer, leads to increased conceptual understanding, resulting in improved performance after peer instruction (Hazel et al. 2013). However, there is an alternative explanation that students do not in fact learn from the discussion, but simply choose the answer most strongly supported by neighbors they perceive to be knowledgeable (Smith et al. 2009). Thus, there is evidence to support both the idea that students are capable of constructing their own knowledge through peer discussion and that certain instructional practices positively affect student attitudes toward discussion (Knight et al. 2013).

7.2.3 Characteristics, Advantages, and Disadvantages of Clickers

In general, the clicker is a small transmitter that is similar to a television remote control, which allows students to quickly answer questions presented in class. Normally, a clicker system consists of three components: (a) clickers: wireless handheld transmitters that resemble small, TV remote controls; (b) receiver: a transportable device that receives signals from the clickers; and (c) software: an application installed on the instructor's computer to record, display, and manage student responses and data. Recently, with the development of ubiquitous computing and network technology, clickers can also be smart phones, tablets, or other kinds of mobile devices with the support of clicker APPs. Herrmann et al. (2012) reported that Dr. Nathaniel Grove from Chemistry Department, University of North Carolina Wilmington, submitted a proposal to the National Science Foundation (NSF) for the development of an interactive personal response system, uRespond. The proposal called for uRespond to include the student client, built on the iPad, a question authoring and class management tool and a central database for storing questions and answers.

Although the morphological characteristics of clickers are changing with technology development, the functions are almost the same as before. When the students answer the questions, the clickers' codes appear on-screen and students know that their responses have been recorded. A computer or tablet, acting as a server, collects and summarizes the feedback and projects the results automatically in the form of a pie chart, a radar map, or a histogram. The feedback from the students can be anonymous since it is linked to the server through the clicker with specific unit ID so that only the teacher himself/herself knows which student gives a correct responses and which gives an incorrect one. Because of the rich interaction between students and teachers, the clickers are frequently used to speed up didactic lectures when teachers teach adults and/or active learners.

Many researchers have suggested that clickers provide significant benefits to both the teacher and students (Bergtrom 2006; Bullock et al. 2002; Simpson and Oliver 2007). Clickers provide immediate feedback about the student's learning process for the teacher, which allow him/her to gauge the overall comprehension of the concepts involved in the material. In addition, clickers are effective tools to enhance students' engagement, to promote interactions among students, to provide immediate feedback on their understanding of the lessons for the teacher, and to facilitate the active participation of students in the learning process by discussing the answers given to the questions. Moreover, the clickers integrate "game approach" that may engage students more than lecture methods (Martyn 2007). These characteristics of clickers encourage the development of student–teacher relationship and lead students to perceive the activity as entertaining, which, in turn, increases their willingness to participate in the class (Caldwell 2007). In short, clickers improve student's understanding of complex subjects, individual progress, and comprehension and teacher awareness of learning problems (Caldwell 2007; Knight and Wood 2005).

In spite of these significant benefits of clickers, some disadvantages should also be noted. First, although the price is decreasing, the cost of clickers is still too high to adopt for some educational institutions, especially the mobile devices with clicker APPs, which becomes a barrier to equip the class and integrate them into the learning process. Second, for all the ease of using the technology and the benefits that they provide, faculty members may be reluctant to introduce new technologies in class and may perceive high costs in terms of time and effort investments

(Kay and LeSage 2009). Third, similar to other advanced technologies, technical issues of clickers may cause frustration and unsatisfactory situations. Overall, the significant benefits identified lead us to expect a positive effect of clickers on student learning performance.

Clickers support various classroom strategies, including the following six proposed by Zhu (2007): (1) assessing students' prior knowledge and identifying misconceptions before introducing a new subject; (2) checking students' understanding of new material; (3) using peer instruction and other active learning strategies; (4) starting class discussion on difficult topics; (5) administering tests and quizzes during lecture; and (6) recording class attendance and participation.

However, clickers also have their disadvantages. D'Inverno et al. (2003) observed that clicker questions are often merely stuck to the beginning or end of lectures. Even educators trying to empirically show the effectiveness of clickers have used them only "very minimally" (Morling et al. 2008; Poirier and Feldman 2007). Further, different types of classes have different goals. Large and introductory courses generally present breadth of material, giving students a taste of the wide range of topics within a discipline. Smaller and upper-year courses give depth, going into much more details of specific areas within a discipline. Laboratory courses typically propose to instill methodological skills. Therefore, the methods of effective clicker use will probably change depending on the purpose of the course. As Wood (2004) pointed out, clickers can be used skillfully or clumsily, creatively, or destructively. It is hoped that instructor will see that clickers can be effectively integrated into their own classrooms and that researchers can use some of the ideas in this paper as a guide to testing the effectiveness of different forms of clicker use.

7.3 Research Design

To compare the effects of two strategies that support active learning, an experiment was designed for a secondary biology course. The students who attend the biology course were invited to participate in this experiment. The students' learning achievements, learning interests, and learning attitudes were collected after learning with the particular subjects of biology during 4 weeks. Cognitive levels were determined by analyzing participants' reflection reports. The research design of this study is described in detail below.

7.3.1 Participants

The participants included eighty-eight students (all females) from four classes of a leading secondary school for girls in Beijing, China. These participants attended the biology course taught by the same teacher and had prior learning experience using clicker and peer discussion. This course was for eighth-grade students in secondary school, and participants were randomly assigned into two groups. The 43 participants involved in the experiment group were taught by clickers along with lecture. The 45 participants assigned to the control group were assigned to learn with peer discussion along with lecture. All the participants were taught by the same teacher with the same learning content, schedule, presentation slides, and questions. Their ages ranged from 13 to 14 years old. Four lessons were selected in a biology course with the duration of each class about 45 min. The teaching materials were based on "Biological Heredity and Variation." This subject focused on how biological evolution and variation works. After finishing the course, the participants could understand biological heredity, biological variation, and genetic diseases and eugenics.

7.3.2 Data Collection Instruments

In this study, four data collection instruments were a pretest and a posttest for the learning achievements, a questionnaire for measuring the learning interests and attitudes of the group regarding the biology course and a learners' cognitive level checklist for surveying the cognitive levels of students according to their reflection reports.

In terms of learning achievements, a pretest and a posttest to evaluate the learning effectiveness of students were compiled by two experienced teachers who had taught this course for more than 10 years. Aiming to ensure that the two groups of students had an equivalent basic prior knowledge of biology course content, the pretest consisted of forty multiple-choice items with a perfect score of 100 for evaluating the students' prior knowledge. The posttest consisted of forty multiple-choice questions asking participants to select the correct answers. The questions of the posttest were totally different to those of the pretest, and the posttest was developed to evaluate the participants' knowledge of new learning contents. The perfect score of the posttest was 100 too.

In addition, the learning interests and learning attitudes questionnaire was modified from that developed by Hwang and Chang (2011). It consisted of 11 questionnaire items for the "learning interests" and 7 items for "learning attitudes" on a five-point Likert scale, where "5" represented "strongly agree" and "1" represented "strongly disagree" (see Table 7.1). The Cronbach's alpha value of the learning interests and learning attitudes were 0.88 and 0.79, respectively, showing good reliability in internal consistency.

Furthermore, learners' cognitive level checklist was revised from that developed by Yen et al. (2012). Learners presented their reflection reports after completing the experiment. The revised Bloom's taxonomy (Krathwohl 2002) was used to divide the content of learning into the six dimensions, such as remembering,

Table 7.1 Questionnaire items for learning interest and attitudes

Learning interests
1. Biology courses are interesting
2. Learning more about biology is interesting
3. Observing different living things is interesting
4. Visiting people in the community to obtain biology knowledge is interesting
5. Learning with peers in the biology course is interesting
6. It is interesting to answer those questions while learning in the field during the biology course
7. I always look forward to taking the biology course and prepare for it before class
8. The teacher's instructions in the biology course have attracted my attention
9. Anything concerning biology is always interesting to me
10. The biology course is more interesting to me in comparison with other courses
11. Other courses do not attract me as much as the biology course
Learning attitudes
1. The biology course is valuable and worth studying
2. It is worth learning those things about biology
3. It is worth learning the local biology well
4. It is important to learn more about biology, including observing and learning different living things
5. It is important to know the biological knowledge related to our life
6. I will actively search for more information and learn about biology
7. It is important for everyone to take the biology course

understanding, applying, analyzing, evaluating, and creating. Two experienced biology course instructors then evaluated the cognitive level presented in learners' reflection reports based on Table 7.2. The evaluation consistency for the two instructors had a correlation coefficient of 0.852 ($r = 0.852$). The total score of each dimension was 5. A higher score indicated that the report content had better cognitive processes in that dimension.

7.3.3 Procedures

The data collection sessions were conducted at roughly four-week intervals. The participants in each group were administered the pretest at the beginning of this experiment. At the four-week intervals, the participants were taught the same learning content about biological heredity and variation. Each group took four classes of teaching in the four-week intervals, and the time for teaching in each class was approximately 45 min. After four-week intervals, they filled out the measurement for learning interests, learning attitudes, and the posttest for evaluating their learning achievements. Also, each student was invited to hand in their own reflection report at the end of the course.

Table 7.2 Cognitive level evaluation for instructor reference

Learning goals of the project tasks	Cognitive levels					
	R	U	Ap	An	E	C
1. *Biological heredity*						
1.1 Able to understand the concept of heredity correctly	☐	☐				
1.2 Able to describe heredity of human correctly	☐	☐	☐			
1.3 Able to analyze the differences of dominant and recessive traits correctly		☐	☐	☐		
1.4 Able to explain genes control traits of mouse				☐	☐	
1.5 Able to understanding the importance of Human Genome Project (HGP)	☐	☐				
2. *Biological variation*						
2.1 Able to understand the concept of variation correctly	☐	☐	☐			
2.2 Able to give some examples of variation correctly	☐	☐	☐			
2.3 Able to analyze the differences between favorable and negative variation correctly				☐	☐	☐
2.4 Able to understanding the significance of variation	☐	☐				
3. *The genetic diseases and eugenics*						
3.1 Able to analyze of cause of genetic disease and genetic patterns correctly			☐	☐		
3.2 Able to calculate recurrence risk of future generations correctly				☐	☐	☐
3.3 Propose strategy and method of controlling genetic disease				☐	☐	☐
3.4 Recommendations on the choice of spouse, family planning advice and guidance				☐	☐	☐

7.4 Results

7.4.1 Student Learning Achievements

Before the experiment, the two groups took a pretest to ensure that they had equal abilities in this subject before the experiment. The mean and standard deviation of the pretest were 77.58 and 9.53 for the experimental group, and 76.31 and 9.38 for the control group. The *t* test result showed that the two groups did not differ significantly ($t = 0.63$, $p > 0.05$), as is shown in Table 7.3. That is, the two groups of students had statistically equivalent abilities before learning the subject unit.

After participating in the experiment, the two groups of the students took a posttest. The ANCOVA result shows that there was no significant difference in students' learning achievements between the experiment group and control group ($F = 0.024$, $p > 0.05$), as shown in Table 7.4. The mean score of the experimental group is 82.74, a little higher than that of the control group, 78.93. In addition, the

Table 7.3 The *t* test result of the pretest scores of the two groups

Variable	Group	N	Mean	S.D.	*t*
Pretest	Experiment group	43	77.58	9.53	0.52
	Control group	45	76.31	9.38	

Table 7.4 Descriptive data and ANCOVA of the posttest results

Variable	Group	N	Mean	S.D.	Std. error.	*F* value	*d*
Posttest	Experiment group	43	82.74	7.697	1.17	0.024	0.52
	Control group	45	78.93	7.340	1.094		

effect size *d* was computed to measure the strength of the *t* test result. In Cohen's definition, "$d = 0.2$" indicates "small" effect size; "$d = 0.5$" represents "medium" effect size, and "$d = 0.8$" means "large" effect size (Cohen 1988). In this study, the *d* value of 0.52 indicates a medium to large effect size, implying that clicker is helpful to students in improving their learning achievements compared to peer discussion.

7.4.2 Learning Interests and Learning Attitudes

Table 7.5 shows the *t* test results of the pre- and post-questionnaire scores of the students in the two groups. It is found that after participating in the learning activities, the control group students' learning interests in the biology course were significantly improved ($t = -3.187$, $p < 0.05$), while their learning attitudes did not show significant improvement. On the other hand, the pre- and post-questionnaire scores of the students in the experimental group show quite positive results. Similarly, from the *t* test results given in Table 7.5, it can be seen that, after participating in the experiment, the students had extremely significant improvements in learning interests ($t = -3.378$, $p < 0.01$). That is, the clicker along with lecture improved students' interests toward learning the biology course.

7.4.3 Cognitive Levels

Two senior biology course instructors evaluated learner reflection reports using a cognitive level checklist. The score of each dimension was analyzed by *t* test. Table 7.6 shows the average, standard deviation *t* value, and *p* value of learner report contents. Table 7.6 shows that, in the six cognitive processes presented in learner reflection reports, different groups exhibited significant differences in the dimensions of remembering ($t = 2.860$, $p = 0.005$), evaluating ($t = -5.037$, $p = 0.000$) and creating ($t = -4.881$, $p = 0.000$). A comparison of the average

Table 7.5 The *t* test result of the pre-and post-questionnaire scores of the two groups

Group	Dimension		N	Mean	S.D.	*t*
Experiment group	Learning interests	Pretest	43	38.09	3.94	−3.378**
		Posttest	43	40.91	3.79	
	Learning attitudes	Pretest	43	25.95	3.21	−2.365
		Posttest	43	27.84	4.12	
Control group	Learning interests	Pretest	45	34.29	4.16	−3.187*
		Posttest	45	37.22	4.56	
	Learning attitudes	Pretest	45	23.84	1.92	−1.863
		Posttest	45	24.80	2.86	

*$p < 0.05$, **$p < 0.01$

Table 7.6 The *t* test results of cognitive levels evaluated from learner report contents

Cognitive levels	Group	N	M	S.D.	*t*	*p*
Remembering	Experiment group	43	4.49	0.471	2.860	0.05*
	Control group	45	4.14	0.640		
Understanding	Experiment group	43	3.30	0.608	0.016	0.987
	Control group	45	3.30	0.770		
Applying	Experiment group	43	3.56	0.625	1.828	0.071
	Control group	45	3.34	0.501		
Analyzing	Experiment group	43	4.44	0.541	1.838	0.700
	Control group	45	4.25	0.422		
Evaluating	Experiment group	43	2.27	0.282	−5.037	0.000***
	Control group	45	2.96	0.865		
Creating	Experiment group	43	2.47	0.285	−4.881	0.000***
	Control group	45	2.82	0.359		

*$p < 0.05$, ***$p < 0.01$

scores shows that (1) in the dimension of remembering, the average score of the experiment group (4.49) was significantly higher than that of the control group (4.14), (2) in the dimension of evaluating, the average score of the control group (2.96) was significantly higher than that of the experimental group (2.27), and (3) in the dimension of creating, the average score of the control group (2.82) was also significantly higher than that of the experimental group (2.47).

7.5 Discussion and Conclusions

The purpose of this study was to compare the effects of two different strategies (clicker and peer discussion) that support active learning on learning performance of secondary students. From the reports of previous studies, clickers have been recognized as an effective approach that have a positive impact on students' learning achievements, classroom engagement, or learning interests, comparing to lecture approach (Addison et al. 2009; Stowell and Nelson 2007). Therefore, it is interesting to examine whether it is the clickers themselves or the more general active learning strategies that influences learning performance.

In this study, an experiment was conducted to evaluate the effects of two different strategies on students' learning achievements, interests, attitudes, and cognitive levels in a biology course. It should be noted that the students in both groups were taught by the same teacher with same learning content, schedule, presentation slides, and questions. The major difference between the two groups was that the students in the experimental group were taught with clicker along with lecture, while those in the control group were taught with peer discussion along with lecture. The quantitative analysis found that there was no significant difference between the two groups in students' learning achievements. Consequently, the effects of the two approaches on learning achievements of both groups were at the same level.

In terms of learning interests, the results were greater for the experimental group than those for the control group. This result revealed that whatever clicker or peer discussion could attract student's attention in the classroom learning. The students' reflections reveals that the students prefer to use clickers than peer discussion because the student with clickers could express their views in anonymity. This finding conforms to those of previous studies that anonymity of clicker appears to increase students' responses (Beekes 2006; Hunsinger et al. 2008; Trees and Jackson 2007). Moreover, there was no significant difference in learning attitudes of the students. A possible reason is that all the participants were from an elite secondary school. Normally, this school only enrolls student with high learning abilities or good attitudes toward learning.

The results of qualitative analysis, using the revised Bloom's taxonomy, indicated that students in the experimental group using clicker performed better than those in the control group in terms of the cognitive levels of remembering which belong to lower levels of cognitive levels. A possible explanation for the results may be that the experimental group using clicker spent more time on slides of the teacher. Moreover, the in-class test by clicker provided more chances for the students to remember what they have been taught before. On the contrary, students in the control group using peer discussion performed better than those in the experimental group in terms of the cognitive levels of evaluating and creating which belong to higher levels of cognitive levels. A possible explanation for the results may be that peer discussion allows very deep, meaningful processing, as each group member adds potentially new and different points to the discussion. This finding conforms to those of previous studies that peer discussion significantly increases scores on concept quizzes (Smith et al. 2009).

Although the control group demonstrated higher cognitive level in the dimensions of evaluating and creating, these benefits did not contribute to students' learning achievements. The possible explanation is that the learning achievements' test is a kind of summative evaluation, which takes the ultimate effectiveness of the overall process in this course into account. Therefore, the learning achievements' test cannot focuses on some special stages, particularly in the high cognitive levels.

There are some limitations in this study. First, the study only involved one instructor with one course. Although the use of one instructor eliminates the potential for a difference of instructor effect that could have occurred if multiple instructors were involved, it opens the door to instructor bias. Without some type of comparison mechanism in place, such as videotaping each lecture or having an individual monitor for all lectures, any differences would be difficult to identify. A second limitation lies in that the pilot school. As one of the best secondary schools in Beijing, it enrolled the elite students with high learning abilities and attitudes toward learning. Thus, the learning interests and learning attitudes of the students may be different from the students in the ordinary secondary schools. A final limitation to the study involves the limited scope of the sample. The sample included only female students.

Future studies may examine whether the findings of this study can be replicated with female students in others courses. Additionally, the sample consisted of traditional students who are familiar with learning and using new forms of technology. Since the number of elder and nontraditional students is increasing, future research may want to examine whether the same results can be found with nontraditional students.

In summary, this study provides evidence to support that there is no significant difference in students' learning achievements between the use of clickers or peer discussion to support active learning. Moreover, clickers contributed more to low cognitive levels, while peer discussion contributed more to high cognitive levels. Although the experimental group shows greater learning interests than the control group, there was no significant difference in terms of the learning attitudes in between the two groups. Lacking statistically significant results in learning achievements in this study, the data of learning attitudes and learning interests still show that students perceive value in the use of clickers. Therefore, it is recommended that clickers be used alongside with peer discussion to enhance active learning in future classes.

References

Abdu-Raheem, B. O. (2010). Relative effects of problem-solving and discussion methods on secondary school students' achievements in Social Studies. Unpublished Ph.D thesis, University of Ado-Ekiti, Ado-Ekiti.

Addison, S., Wright, A., & Milner, R. (2009). Using clickers to improve student engagement and performance in an introductory biochemistry class. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology, 37*(2), 84–91.

Auster, E. R., & Wylie, K. K. (2006). Creating active learning in the classroom: A systematic approach. *Journal of Management Education, 30*, 333–354.

Beatty, I. D. (2005). Transforming student learning with classroom communication systems. *arXiv preprint physics/0508129.*

Beatty, I. D., Grace, W. J., Leonard, W. J., & Dufense, R. J. (2006). Designing effective questions for classroom response systems teaching. *American Journal of Physics, 74*(1), 31–39.

Beekes, W. (2006). The 'millionaire' method for encouraging participation. *Active Learning for Higher Education, 7*, 25–36. Retrieved on September 10, 2013 from http://alh.sagepub.com/cgi/reprint/7/1/25

Bergtrom, G. (2006). Clicker sets as learning objects. *Interdisciplinary Journal of Knowledge and Learning Objects, 2*. Available in http://ijklo.org/Volume2/v2p105-110Bergtrom.pdf

Bonwell, C., & Eison, J. (1991). Active learning: Creating excitement in the classroom, ASHEERIC Higher Education Report No. 1. Washington, DC: The George Washington University, School of Education and Higher Education.

Brady, M., Seli, H., & Rosenthal, J. (2013). "Clickers" and metacognition: A quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback devices. *Computers & Education, 65*(1), 56–63.

Bullock, D. W., LaBella, V. P., Clingan, T., Ding, Z., Stewart, G., & Thibado, P. M. (2002). Enhancing the student-instructor interaction frequency. *The Physics Teacher, 40*, 535–541.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and beast-practice tips. *CBE Life Sciences Education, 6*(1), 9–20.

Chen, J. C., Whittinghill, D. C., & Kadlowec, J. A. (2010). Classes that click: Fast, rich feedback to enhance students' learning and satisfaction. *Journal of Engineering Education, 99*(2), 158–169.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

D'Inverno, R., Davis, H., & White, S. (2003). Using a personal response system for promoting student interaction. *Teaching Mathematics and Its Applications, 22*, 163–169.

Duncan, D. (2006). Clickers: A new teaching aid with exceptional promise. *Astronomy Education Review, 5*, 70.

Hazel, S. J., Heberle, N., McEwen, M. M., & Adams, K. (2013). Team-based learning increases active engagement and enhances development of teamwork and communication skills in a first-year course for veterinary and animal science undergraduates. *Journal of Veterinary Medical Education, 40*(4), 333–341.

Herrmann, A., Reinicke, B., Vetter, R., Yaylacicegi, U., & Grove, N. (2012). uRespond: A classroom response system on the iPad. *Annals of the Master of Science in Computer Science and Information Systems at UNC Wilmington, 6*(1).

Homme, J., Asay, G., & Morgenstern, B. (2004). Utilisation of an audience response system. *Medical Education, 38*(5), 575.

Hunsinger, M., Poirier, C. R., & Feldman, R. S. (2008). The roles of personality and class size in student attitudes toward individual response technology. *Computers in Human Behavior, 24*, 2792–2798 (Jones, C., Connolly, M., Gear, A., & Read, M. (2001). Group).

Hwang, G.-J., & Chang, H.-F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023–1031.

Kay, R.-H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: a review of the literature. *Computers & Education, 53*, 819–827.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education, 4*(4), 298–310.

Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE-Life Sciences Education, 12*(4), 645–654.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218.

Lee, J. B., & Bainum, C. K. (2006, April). Do clickers depersonalize the classroom? An evaluation by shy students. In *86th Annual Convention of the Western Psychological Association, Palm Springs, CA.*

Martyn, M. (2007). Clickers in the classroom: An active learning approach. *Educause Quarterly, 30*(2), 71.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology, 34*, 51–57.

McKeachie, W. J., & Svinicki, M. (2006). *Teaching tips: Strategies, research, and theory for College and University teachers*. Belmont, CA: Wadsworth.

Meyers, C., & Jones, T. B. (1993). *Promoting active learning, strategies for College classroom*. San Francisco: Jossey-Bass Publishers.

Morling, B., McAuliffe, M., Cohen, L., & D'Lorenzo, T. (2008). Efficacy of personal response systems ("clickers") in large, introductory psychology classes. *Teaching of Psychology, 35*, 45–50.

Poirier, C. R., & Feldman, R. S. (2007). Promoting active learning using individual response technology in large introductory psychology classes. *Teaching of Psychology, 34*, 194–196.

Preszler, R. W., Dawe, A., Shuster, C. B., & Shuster, M. (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE Life Sciences Education, 6*(1), 29–41.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223–231.

Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *American Journal of Distance Education, 17*(2), 77–98.

Simpson, V., & Oliver, M. (2007). Electronic voting systems for lectures then and now: A comparison of research and practice. *Australasian Journal of Educational Technology, 23*(2), 187–208.

Smith, M., Wood, W., Adams, W., Wieman, C., Knight, J., Guild, N., et al. (2009). Why peer discussion improves student performance on in-class concept questions. *Science, 323*, 122–124.

Stowell, J. R., & Nelson, J. M. (2007). Benefits of electronic audience response systems on student participation, learning, and emotion. *Teaching of Psychology, 34*(4), 253–258.

Tai, L. Y. (2011). Integrating the technological dimension into teaching and learning: A sociocultural perspective. Retrieved July 31, 2013, from http://www.cdtl.nus.edu.sg/tech-in-he/pdf/Section2-Article6.pdf

Trees, A. R., & Jackson, M. H. (2007). The learning environment in clicker classrooms: Student processes of learning and involvement in large university-level courses using student response systems. *Learning, Media and Technology, 32*, 21–40.

Wood, W. B. (2004). Clickers: A teaching gimmick that works. *Developmental Cell, 7*, 796–798.

Yen, J.-C., Lee, C.-Y., & Chen, I.-J. (2012). The effects of image-based concept mapping on the learning outcomes and cognitive processes of mobile learners. *British Journal of Educational Technology, 43*(2), 307–320.

Zhu, E. (2007). Teaching with clickers. *Center for Research on Learning and Teaching Occasional Papers, 22*, 1–8.

Part III
Smart Pedagogy

Chapter 8
Smart Learner Support Through Semi-automatic Feedback

Paul Libbrecht, Wolfgang Müller and Sandra Rebholz

Abstract Learning tools that produce automated feedback are becoming commodity, from multiple-choice questions to intelligent tutoring systems, and from direct manipulations to exploratory environments. In this paper, we argue how such learning tools can become smart by applying the semi-automatic feedback paradigm where the teacher complements the feedback capabilities of the learning tool. The approach employs analytics as a central awareness mechanism for teacher to provide guidance in a way that is most relevant to the past usage of the learning tool, including what it provided as feedback. The SMALA approach we describe is realized as an open-source software which has been evaluated in a number of undergraduate studies, leveraging the default learning management system's architecture of the universities. This software delivers visualizations of the activities at each level of interaction (the group of all users, the group of users in a classroom, the individual learner). The different levels support the teacher in adjusting his or her strategy and respond to individual requests.

Keywords Semi-automatic · Feedback · Formative · Assessment · Teaching analytics

8.1 Research Motivation and Contribution

The usage of learning tools with automated feedback has been the subject of multiple research, be it with intelligent tutoring systems, with explorative tools, or with even simpler training systems. All are praised to support the differentiations

P. Libbrecht (✉)
Computer Science, University of Education Weingarten, Weingarten, Germany
e-mail: libbrecht@ph-weingarten.de

W. Müller (✉) · S. Rebholz (✉)
Media Education and Visualization Group, University of Education Weingarten, Weingarten, Germany
e-mail: mueller@md-phw.de

S. Rebholz
e-mail: rebholz@md-phw.de

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_8

between the knowledge and learning processes of each learner as well as to provide feedback in arbitrary contexts of work.

However, all of them suffer from a common issue: The quality and relevance of the feedback they can provide is limited to what designers of the software considered thinkable. It has, thus, been natural to consider the teacher as a tutor that coaches learners in the usage of the learning tool, being able to advise at the right time.

In this paper, we contend that learning tools' automatic feedback can become smart by employing the semi-automatic feedback paradigm whereby the teacher is able to complement the feedback generated by the learning tool by a relevant feedback. To this end, we describe the model of a system that supports the teacher in analyzing the learner's process and produce a feedback that is relevant to the learner's context of work. This support is a form of integrated learning analytics.

The model is illustrated by an implemented system containing two learning tools and integrated with two different learning management systems. The research presented here reports on learning tools and evaluations in the studies for pre-service teachers in a project called SAiL-M.[1]

In this environment, as in most higher education environments, the learners are considered mostly responsible of their learning process but can be supported by individual feedback relevant to their work.

8.2 Literature Review

8.2.1 Smart Learning Environments

The term "*smart learning environments*" has been introduced recently in the field of learning and teaching to describe "… a third pervasive and significant revolution in instruction" (Dodds and Fletcher 2004). However, similar to the beginnings in the field of intelligent tutoring systems, to date, there is no completely agreed understanding on this term. Dodds and Fletcher define them as "… real-time adjustment of instructional content, sequence, scope, difficulty, and style to meet the needs of individuals suggests" (Dodds and Fletcher 2004). On the other hand, smart learning environments are also often understood as an improvement of physical environments with novel technologies to provide a smart, interactive classroom with increased interactivity, personalized learning, efficient classroom management, and better student monitoring (Yesner 2012). Last but not least, smart learning environments are also often related to ambient technologies, describing learning environments, which exploit new technologies and approaches, such as ubiquitous and mobile learning, to support people in their daily lives in a proactive yet unobtrusive way (Mikulecký 2012; Buchem and Pérez-Sanagustín 2013).

[1] The SAiL-M project has been funded by the German ministry for research and education. See http://sail-m.de/.

In the following, we would like to understand smart learning environments as systems that apply novel approaches and methods on the levels of learning design and instruction, learning management and organization, and technology to create a context for learning that provides learners with opportunities for individualized learning and reflection in a motivating way and that allow teachers to facilitate learning, providing scaffolding and inspiration based on the learner's needs and a careful observation of her learning activities.

Therefore, we also would like to point out that in our understanding approaches in the direction of smart learning environments cannot be restricted onto the technological level, only.

8.2.2 Intelligent Tutoring Systems

Sleeman and Brown first coined the term *intelligent tutoring systems* (ITS) to describe the problem-solving steps of the student through the use of a detailed cognitive model of the domain (Anderson and Pelletier 1991), which distinguished from previous computer-aided instruction (CAI) systems (Sleeman 1982). It is therefore interesting to note that in the beginning, ITS did not denote a well-understood principle or a commonly agreed approach. Current definitions describe them as learning systems that give feedback and hints on each step (VanLehn 2011). Another perception is the one as a model-tracing tutor, where the machine takes the role of a human tutor and follows the inputs of the learner.

While intelligent tutoring systems failed for quite a while to have a real impact on education and training in the world (see Corbett et al. 1997a, b), today there exist a number of success stories, and ITS have been applied successfully in some domains. At present, systems such as ASSISTments (Feng et al. 2009; Worcester Polytechnic Institute 2013), a free Web-based service supporting math classes in US American schools, are being used by hundreds of students each year on a regular basis since 2004 and did prove its benefits in several studies (e.g., Heffernan et al. 2012). Also, technologies simplifying the development of intelligent tutoring systems even for non-programmers were made available (e.g., Koedinger et al. 2003, 2004), allowing for a faster development of such systems for different application areas.

Still intelligent tutoring systems did not fulfill all promises that were made earlier. Some reasons for a lack of penetration of *intelligent tutoring systems* were already stated over 15 years ago (Corbett et al. 1997a, b), but in principle, they still seem to hold today. Intelligent tutoring systems are still expensive to develop, and in practice, the development requires sufficient resources with respect to programming, despite the mentioned initiatives to overcome these limitations. For this reason, still many initiatives in this area are initiated in the fields of computer science and artificial intelligence (AI) in specific, and they often seem to focus on the deployment and improvement of interesting AI algorithms, rather than emphasizing the educational perspective and trying to enhance the cost/benefit trade-off regarding educational effectiveness (Corbett et al. 1997a, b). This may also have lead to relative low reputation of the value of intelligent tutoring systems, though VanLehn (2011) was able to show up in a recent study that the effect size of intelligent tutoring systems is nearly as effective as human tutoring.

Paradigms from computer-aided instruction targeted to cost reduction and enhanced objectivity in teaching can still be attributed as driving forces also for developments in the area of intelligent tutoring systems. As such, the idea to have an intelligent tutoring system replacing the teacher is still predominant. While this has been criticized early and requests for supportive cognitive tools in the service of explicit pedagogical goals, supporting both, learner and teachers were raised (e.g., Reusser 1993), this focus still seems to prevail in most ITS applications today.

The approach presented in this article distinguishes from the typical concepts found in intelligent tutoring insofar that it follows more closely the recommendations of Reusser (1993) to focus on a better integration into school routine and provide a smart support also for the teacher and tutor.

8.2.3 Assessment and Feedback

Smart learning environments and intelligent tutoring systems both relate to approaches in the field of assessment. Assessment usually has one of the following two purposes:

- *Judgment*. In this type of assessment—usually denoted as *summative assessment*—the goal is to decide whether the learner has passed a course and at what level or grade. It is mostly an *assessment of learning* and used to measure students' understanding of a specific topic (Ainsworth and Viegut 2006).
- *Development*. Assessment in this context is denoted as *formative assessment*. Here, assessment results are used by teachers to analyze the students' concepts and levels of understanding with the intention to adapt teaching according to the students' needs and to provide adequate feedback, and by learners who can evaluate their advances based on this feedback. This is an *assessment for learning*, and the results are not to be used to grade students' work (Ainsworth and Viegut 2006).

Formative assessment and feedback have been identified as important factors in teaching and learning with a high effect-size (e.g., Hattie 2008), and adequate feedback plays a crucial role in formative assessment (Bescherer et al. 2009; Brown 2004). Shute defines formative feedback as "… information communicated to the learner that is intended to modify the learner's thinking or behavior for the purpose of improving learning" (Shute 2008). Brown (2004) states that formative feedback "needs to be detailed, comprehensive, meaningful to the individual, fair, challenging and supportive, which is a tough task for busy academics" (Brown 2004). Adequate feedback is agreed to represent an important factor in the support of learners, and it was found to have a potential high impact on learning, though some studies revealed that inadequate feedback may also obstruct learning (Kulik and Kulik 1988; Anderson et al. 1990). Besides several other aspects, timing was often named as an important factor for delivering effective feedback, and immediate feedback proved to be more effective than delayed one in several studies (Anderson et al. 1990; Butler et al. 2007; Shute 2008; Singh et al. 2011). However, the results are inconsistent, and Hattie (2008) identified only low effect-sizes regarding the effect of timing in his meta-studies. A possible explanation is that it depends on level (task level versus process level) and task difficulty whether immediate or delayed feedback is beneficial (Hattie and Timberley 2007).

Hattie also points out the importance of feedback not only for the learner, but also for the teacher, to synchronize learning and teaching and to make both more effective (Hattie 2008; Chap. 9 The contributions from teaching approaches - I: Feedback). This aspect relates to the aforementioned strong connection of feedback to assessment, highlighting however the information gain at the side of the teacher in feedback processes. The approach presented in the paper in particular takes on this aspect, providing a framework that allows students to receive timely and detailed feedback on specific process steps, but also allowing teachers to "make learning visible" (Hattie 2008) and to access detailed information on the learning processes of classes and individuals.

In technology-enhanced learning, a clear focus was on the automation of providing feedback to learners in the last years, leading to solutions strongly connected to the fields of artificial intelligence and intelligent tutoring systems. Lately, a different approach and philosophy was proposed, where technology gets the role to assist teachers in providing feedback in an automatic or semi-automatic way, encompassing all cases where feedback may be provided following standardized schema (Müller et al. 2006; Bescherer et al. 2011). This disburdens the teacher from managing assessments and feedback in these standard cases, allowing her to focus on interesting cases and situations where very specific feedback is required. Corresponding approaches lead to completely different system architectures, where the teacher plays an important role, and assessment and feedback are provided in a hybrid approach and a semi-automatic way. The work presented in this chapter closely follows this approach.

8.2.4 Learning Analytics and Teaching Analytics

Computer-based learning tools provide the benefit that all sorts of digitally available learning data can be collected and analyzed. Corresponding approaches fall in the field of *learning analytics*. There have been a number of proposals for defining learning analytics, which to some extent take different objectives and only partially overlap (Siemens 2011). We follow Rebholz et al. (2012) and we relate learning analytics to approaches and technologies targeted to allow for analytical reasoning facilitated by visual interfaces employed for teaching or learning.

Objectives are the detection of interesting aspects and patterns in learner and learning data, building hypotheses based on these detected structures, confirming such hypotheses, drawing conclusions, and possibly communicating the results of this analytical process (Rebholz et al. 2012).

Teaching Analytics is sometimes used to denote specific approaches in the field of learning analytics. However, the term is often used inconsistently. For instance, it is used on the one hand to describe a subfield of learning analytics, focusing on the design, development, evaluation, and education of visual analytics methods and tools for teachers in primary, secondary, and tertiary educational settings (NEXT-TELL 2013). On the other hand, Vatrapu et al. (2011) use the term *Teaching Analytics* in the context of an extended view and model, targeting to a collaborative analysis of learning data with teaching experts, visual analytics experts, and design-based research experts.

In learning analytics, the focus is often on the prediction of student performances with respect to learning across a variety of courses and academic programs, and on the identification of at-risk students and the design of educational interventions (e.g., Arnold 2010; Zhang and Almeroth 2010; Essa and Ayad 2012). Far less frequently, learning analytics methods are being applied to monitor student performance on the level of individual tasks and learning processes. Corresponding approaches require a more deliberate design of adequate interfaces. Such interfaces shall support teachers in the effective interactive analysis of learning data and allow them to provide timely, meaningful actionable, customized, and personalized feedback to students (Vatrapu et al. 2011; Rebholz et. al. 2012).

Learning management systems such as the widespread Moodle platform offer elementary forms of user tracking, displaying such information as the start or end of use of learning resources, the visits, or the participation to pre-built activities such as polls. Moodle's tracking, by default, is very detailed and could be used to understand a learner's progress and thus can be considered to play a learning analytics role. However, most tracking views are of tabular nature, which quickly become unmanageable as soon as a large diversity of tracked events may appear, e.g., a diversity of user input or feedback sequence. We contend that such generic tracking systems lack the specificity of learning tools, which allows a teacher or learner to understand the steps of the solution process.

Besides such generic logging mechanisms in learning management systems, some more sophisticated approaches have appeared in learning analytics. LOCO-Analyst (Jovanovic et al. 2007, 2008) represents a logging infrastructure targeted to provide teachers with detailed information on students' learning processes based on their interactions with learning objects. The goal is to generate meaningful feedback for educators responsible for updating and revising course material. While the LOCO-Analyst approach has similarities to the logging functionalities proposed and presented in this chapter, it seems to be restricted to track accesses to different learning objects and does not allow a fine-grained view of the flow of actions of a single user using dynamic learning tools such as those producing feedback. The FORMID project (Guéraud-Cagnat and Cagnat 2006) instead produced monitoring facilities for e-learning, supporting the online support of teachers for students' learning activities in real time. The work presented here differs from the FORMID approach in that such a support is not tight to a learning scenario. As a result, logging mechanisms, user interfaces, and analytical views need more flexibility.

Recording learning activities and making the recordings available for further analysis require suitable logging infrastructures. The emerging set of specifications called Tin Can API (http://tincanapi.com/), the open-source tool sets Contextual Attention Metadata (https://sites.google.com/site/camschema/) and Learning Registry (http://www.learningregistry.org/) all target at capturing learning activities and store them in central repositories. The major drawback of these solutions is the difficulty of doing statistics on the data without dedicated log viewers offering views that are sufficiently representative of the learner activity to understand quickly the solution process and where the problems were met.

8.2.5 Security and Anonymity in Learning

A teacher watching over the shoulder of a student performing an exercise and receiving feedback about her actions is in a situation of common agreement: The student may feel like she could explore or *game the machine*, but he or she will certainly not want to show complete ignorance of the presence of the teacher by acting unreasonably, for example.

Learning activities tracked by a logging system and potentially displayed to a teacher create the same situation. The ignorance of being watched and the surprise of being told by a teacher that a given action could have been better done would trigger a flurry of counter-reactions in the students' minds, which may go as far as refusing to use the learning tool. Such issues about privacy have been quite ignored by the literature in learning analytics as far as we could read.

National regulations in EU countries about the usage of Web sites boil down to prohibit the storage of personal information without consent (e.g., German Federal Data Protection Act (BDSG) §3, §4) and, thus, also requiring the allowance before a detailed tracking of users.

These two limitations thus let us conclude that:

- Learners should be warned about being tracked and about teachers being able to view anonymous logs.
- Learners should be allowed to stop tracking, should they want to play with more freedom.
- Learners that wish to disclose their work sessions as a personal series of action should be able to do so. This way, they enable teachers to understand what they have done.

8.3 Research Design

In the following, we introduce our research design. For this, we first present example scenarios motivating and serving as a guideline for our work. Then, we provide a short overview on the general approach and the research methodology taken.

8.3.1 Motivating Examples

8.3.1.1 Scenario I: From Demonstration to Homework

Our learning situation happens in an undergraduate study with mathematics. Mary, the teacher, has been introducing the topics of mappings, injections, and surjections. After the theory and examples of different types, the lecture had a demonstration of the learning tool *Squiggle-M*, which explores this topic. The week's assignment includes a few exercises to be done using the learning tool such as the exercise displayed in Fig. 8.1.

In the afternoon, Philip is doing his homework and now follows the steps of the learning tool. The first few exercises are easy with the help of the videos. He then attempts to recognize when a graph of a function is injective and faces challenges. Going back and checking definitions helps him a little bit, but he only succeeds in

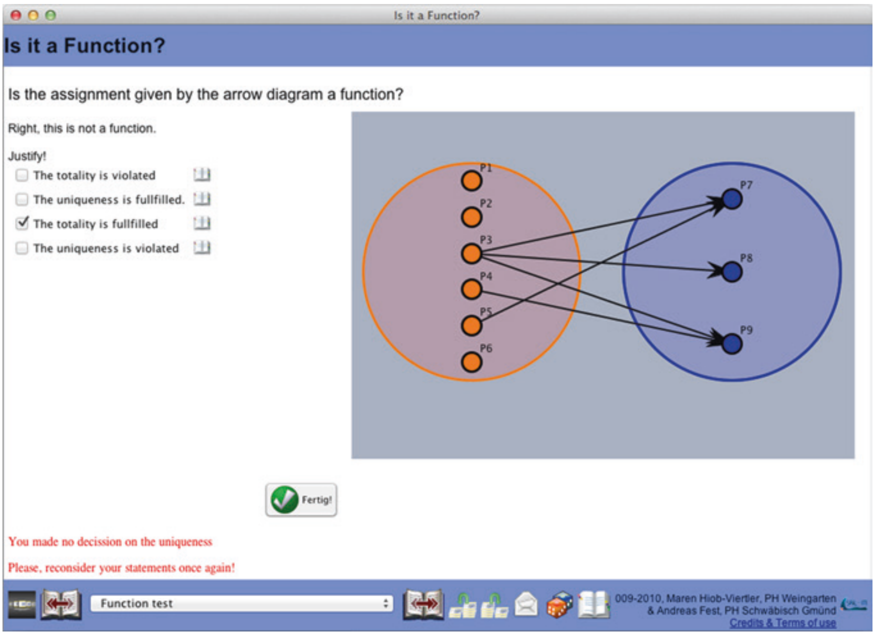

Fig. 8.1 The Squiggle-M learning tool asking if a relation qualifies as a function

the simplest case of an exponential function. He thus wishes to ask his teacher and sees that the problem reports indicate this possibility. A dialog opens, similar to that of Fig. 8.2 where he can input a short description of what he wishes to achieve and a checkbox is there to give the teacher access to Philip's logs to understand what was done.

Mary receives a notification per email. It contains a screenshot and a link to the session that Philip just did. Looking at the log, Mary can analyze what he did, reproduce it herself, and give short hints on the actions: "To evaluate injectivity, it may be useful to click twice the red bullet and lay it on the x-axis, so that you can see if you can have two points that are mapped to one." Based on this feedback, Philip is able to employ the right tool and evaluate injectivity and subjectivity for each of the proposed functions.

8.3.1.2 Scenario II: Encouraging the Tools' Usage

Jonathan is a university teacher educating future mathematics teachers. He wishes to introduce the proper use of proofs by induction, a topic that is well known to create confusion in young students but remains quite important for many proofs of the mathematical knowledge. Thus, he decides that the usage of a learning tool to train such proofs is desirable. ComIn-M (Rebholz and Zimmermann 2013) is such

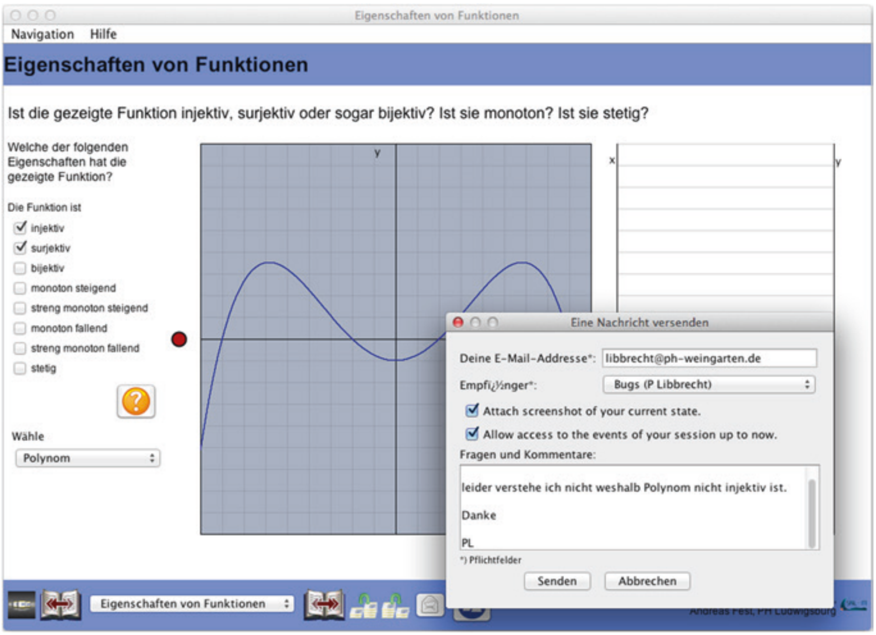

Fig. 8.2 Squiggle-M displaying a function by its graph and a ladder with a dialog of the student's question to the teacher being formulated

Fig. 8.3 Explaining the usage of in interactive learning tool

a learning tool. It can be run on contemporary laptops' and desktops' Web browsers, which also access the learning management system of the university for all students.

Jonathan contacts the editors of the learning tool, which provide him with an online learning *activity*. In there, he can read the instructions to deploy the learning tool within the learning management system: Simply uploading a content package will create an online resource from which students can start the learning tool. He shall make it visible a bit later.

Following the didactical design pattern *Technology on Demand* (Bescherer and Spannagel 2009), Jonathan first presents a few situations of proofs by inductions and its typical errors in class and then introduces the usage of the tool. To do so, he presents the tool and performs one complete exercise with it. Somewhat similarly to the Fig. 8.3, Jonathan is able to present connections between words and graphics of the blackboard and learning tool projection. His approach follows mostly the orchestration[2] *explain-the-screen* (Tabach 2013). At the end of the session, he invites the students to use the tool, demonstrating how it can be started in the learning management system; one of the exercises of this week's assignment is based on the learning tool.

Because exercises are optional, he cannot be sure that the exercises will be performed. A few days later, as he feared, very few students actually attempted the requested exercise from what he can see in the log views in the Fig. 8.4: Only 4 of his 150 students have attempted, and, as he can see in the assessment results table on the right, none have succeeded. In the graphic below, red cells represent wrong solutions, light red cells incomplete solutions, while (the missing) blue cells would represent correct solutions.

For the following exercise session, thus, a plan change is communicated so that students come with their laptops to the university. The objective of Jonathan is to let the students go through as many of the ComIn-M exercises as possible in small groups in front of the laptops keeping eyes wide open to ensure that they are progressing.

[2] The classroom orchestration, explained for example in Tabach (2013), is a description of the didactical configuration of the classroom that is well suited to describe the usage of technological tools.

Fig. 8.4 A summary view to gain overview of the class's usage of the learning tools

During the help session, students are first given a briefing on the mission they are to aim at, assorted with a set of practical and strategic instructions. Most of the rest of the class is spent in the classroom orchestration *monitor-and-guide*, where the teacher comes at each student providing individualized help on demand, answering typical help requests in just a few minutes in an attitude similar to that similar the situation depicted in Fig. 8.5. The teacher's work there generally involves understanding the students' states, what they have done to reach it (which can be shown or told by the students, e.g., a particular type of problem, which keeps being reported by the learning tool) and what they understand to have made these manipulations.

The decision to help can either be following a students' initiative or a teacher's observation. This observation can be over the shoulder or based on some analytics representations.

8.3.2 Derived Research Design

The scenarios described above describe the underlying vision of a research collaboration and project targeted to improve the quality of teaching in early semesters at university level. The project titled SAiL-M (Semi-automatic analysis of individual learning processes in mathematics) focused on the domain of mathematics, although the developed concepts can be generalized to other areas.

Fig. 8.5 Helping students in presence

In this project, a design-based approach in educational research (Bannan-Ritland 2003; Reeves et al. 2005) was followed, starting with the mining and formulation of pedagogical design patterns for activating learning environments for mathematics at university level, the advancement of corresponding scenarios, followed by an adaption of learning tools, which allow for the assessment of learning processes. These tools and learning environments as well as the developed application scenarios were applied and evaluated. Here, the focus was on evaluating the effectiveness of process-oriented feedback with various diagnostic methods. The evaluation also concentrated on the *local impact* of the implemented methods and applied toolsets (Bannan-Ritland 2003), limited to pre- and posttests and not incorporating control groups.

In this chapter, we report selected results from this collaboration and initiative, which focus on the aspects of detailed logging of learners' activities, semi-automatic assessment and feedback, and learning analytics solutions to support teachers in providing adequate support to learners. A general model, which has been developed in the context of the SAiL-M project, underlies the developed methods and toolsets. We introduce this model and the corresponding developments in the following section.

8.4 The System

The scenarios and research design discussed in the previous section led us to a novel model for learning scenarios with enhanced support for learners based on adequate and timely feedback, and a stronger integration of teachers in this process compared to standard intelligent tutoring systems. We describe and discuss this model in some more detail in the first section. Furthermore, we present tools and techniques, which were developed in a research cooperation in Germany implementing this model.

8.4.1 Proposed Approach: The SMALA Model of Smart Learner Support

In this section, we describe a conceptual model, which places the learning tools within an architecture that enables the automatic and semi-automatic feedback.

We call the model the *SMALA model*. Its purpose is to provide relevant feedback on the learning processes that occur in multiple situations of the learners. These situations are, in no particular order:

- Learning in classroom, in a plenum, where the lecture concepts are fresh in memory.
- Learning in the lab, in small groups, or individually, where individual assistance can be provided to specific requests.

- Learning in homework and other rehearsal situations, where the student's liberty to explore potential avenues is greatest.

Each of these situations implies different forms of feedback and different types of reflection on the individual learning process. In the center of it lies the learning tool, which *represents* the domain being learned and offers the necessary manipulatives. The SMALA model complements the automated learning tool by an architecture to support the teachers to provide feedback relevant to the learning process beyond the automated feedback of the learning tool. This complement is ensured by:

- A deployment within the existing learning tools infrastructure of the school: the Learning management system (LMS, the normal place where learning activities are coordinated).
- The recording of traces of the actions of the learning tools in a way that allows sequences of actions to be viewed.
- The display of anonymized traces of the usages of the learning tools to the teachers to obtain an overall impression of their usage.
- The display of identified traces of the usages when the learners explicitly request feedback.

The picture in Fig. 8.6 is a summary of the architecture of this model. It highlights where *analytical processes* happen (where the gears icons appear) and the workflow taken by the teacher to prepare the learning tool so that it is ready to be used and tracked. This includes obtaining a new *activity*, which encompasses the planned set of learning tools' usage, the deployment of the learning tools in the LMS, and the invitation of the students to that place. This enables the students' logs to be attached in the right surrounding and makes them browsable by the teacher: individually when the student asks, globally (i.e., anonymously) otherwise.

In this model the feedback is produced either by the learning tools' automatic assessment, or by the teacher: individually on students' request, or globally, for example, in classrooms.

8.4.2 Tools and Techniques: From Learning Tools to Analysis

In the following, two SMALA-based learning tools are presented in detail: a training tool for proving by mathematical induction and a learning tool for investigating the concept of relations and functions. Both the instructional design and the technological implementation underlying these tools are highlighted. As indicated in the motivating scenarios before, we distinguish between the students' view on the learning tools and the analytics' view for the teachers.

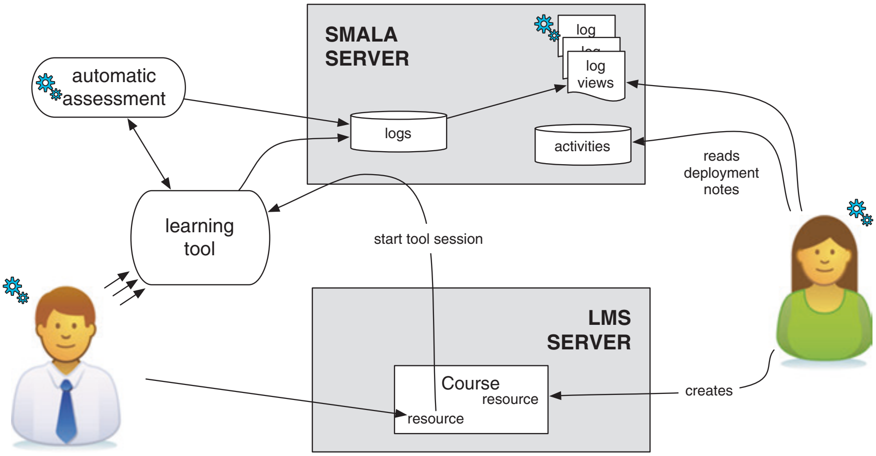

Fig. 8.6 System architecture corresponding to the SMALA model

8.4.2.1 From Learning Design to Learning Tools to Analysis

In order to support learners in introductory math classes at university, we have integrated various Web-based learning tools into the SMALA infrastructure that offer training and advanced investigation opportunities on specific mathematical subject domains. Although the tools are quite different in focus (e.g., discovery-oriented tool versus training tool), they all rely on some common design principles:

Semi-automatic assessment and feedback. Based on the principle of semi-automatic assessment and feedback (Bescherer et al. 2011), all tools provide the learners with immediate feedback on their solutions and, if necessary, enhance the feedback by personal advice from a tutor or teacher. As soon as a learner submits a solution or partial solution, the automatic assessment component analyses the solution for correctness. As part of the analysis, the software checks for typical errors or misconceptions and generates a detailed feedback report based on the findings. If the analysis fails to verify the solution or is not able to detect and classify an erroneous step in the solution, the problem is forwarded to the assigned teacher. Using both the results from the automatic analysis of the learning tool and a recording of all interactions between the learner and the tool, the teacher can reproduce the chosen problem-solving strategy. By doing so, the teacher gets an idea of the individual learning process, and thus, possible misconceptions and errors in reasoning become obvious. In the same way, correct but exceptional solutions are forwarded to the teacher and do not remain unnoticed in the wealth of learner data arising during the tool's usage. This approach supports our main goal of relieving the teacher from repetitive, non-demanding tasks, but involves him or her in the assessment and feedback process when creative understanding, didactic skills, or the domain expertise of the teacher are required.

Step-based tutoring system. According to the classification of tutoring systems suggested by VanLehn (2011), our learning tools are denoted as *step-based tutoring systems*: Learners can enter all intermediate steps that lead to the final solution, and accordingly, they also get feedback on these individual steps. As opposed to *answer-based systems* that only assess the final solution of a problem, our toolset analyses and assesses the whole line of problem-solving steps and aims at reducing the amount of reasoning required between individual interactions with the system. By giving feedback and hints on the level of intermediate steps of the problem-solving procedure, the learners are gradually guided toward generating a correct solution.

Tracing of learning processes. All interactions between the learner and the learning tools are recorded as events by the SMALA logging service. There are three main requirements that are fulfilled by the SMALA logging service: The logging happens transparently (without having the learner notice or feeling disturbed by the event recording), live (as a continuous, real-time stream of events) and by pseudonym. In order to account for data security and privacy issues, we ensured that logging events are not traceable back to the real person that has used the learning tool. However, pseudonyms and session identifiers that are attached to each interaction event enable the reconstruction of learning processes of individual learners, given that the pseudonym is associated to the person.

Analysis of learning processes and learning group performance. SMALA-enabled learning tools use one common infrastructure for storing and analyzing the interaction events that occur during the learning tools' usage. Information provided by the learning tools includes, among other things, the individual steps and solutions entered by the learners, the automatic analysis results, including a success score, and feedback generated by the learning tools, event date, and time, and feedback and hint requests. The events are organized in an expandable hierarchy of objects. All learning data are analyzed in real time and displayed to the teachers in a Web-based user interface on demand. Visualizations, textual lists, as well as table-based representations are used to present the analysis results. It is finally up to the teacher to interpret the automatic analysis results and draw consequences for the subsequent instructional design. Combining technological evaluation with human expertise seems most promising to us when realizing formative assessment strategies.

8.4.2.2 Smart Learner and Teacher Support

By combining these design principles in one system, a technology-based learning infrastructure can be set up that uses as much automatic processing and analysis capabilities as possible, but involves the teacher in the assessment cycle whenever it is necessary to optimally support the learners in their learning processes or to improve and adapt instruction to current needs. This smooth transition between automatic and human activity supports both learners and teachers in a smart way. As opposed to typical ITS systems that only address learners, we would like to emphasize the fact that the SMALA toolset is targeted at both learners and teachers and that both of them shall benefit from technology as much as possible. In the section below, we detail the instruments that allow a smart learner and teacher view.

8.4.2.3 Teacher Preparation

We expect teachers intending to use learning tools with a SMALA toolset to use a learning management system such as Moodle. This enables them to share Web pages, and the learning tools we consider can be run within Web pages. Using this, as well as some LMS-integrating components, an identity of the student can be obtained and thus a pseudonym can be computed.

After having contacted editors of the SMALA server, an *activity* is created which defines the learning tool, access rights, and methods of identification. The teacher's preparation involves reading the deployment instruction and copy and pasting the necessary code to the learning management system. In our project, integration components have been realized for the Moodle and StudIP learning management systems.

8.4.2.4 The Students' View

Using the example of the ComIn-M and Squiggle-M learning software, the students' view on these interactive learning tools is described. After shortly describing the idea underlying each tool, the main characteristics of the user interface and its usage related to semi-automatic assessment and feedback are presented.

Example 1: ComIn-M—Proving by mathematical induction

Idea. The learning tool ComIn-M is a Web-based exercise sheet for training proofs by mathematical induction of elementary arithmetic relations. Students can choose among different summation formulæ that shall be proven by induction. According to the principle of a step-based tutoring system, all intermediate steps leading to the final solution have to be entered. In the case of mathematical induction, this implies that not only the basic procedural steps of the proving process have to be run through, but also very fine-grained steps like individual term transformations for showing equality of expressions. Whenever the learner gets stuck or wishes to get a confirmation that she is on the right track, she can request feedback on the current state of the solution or request help by retrieving hints for the current step. The main focus of ComIn-M is to foster the procedural knowledge of applying proofs by mathematical induction and to offer learning opportunities in homework or rehearsal situations.

Semi-automatic assessment and feedback. As the learner is working through the ComIn-M exercise sheet, she enters the solution step by step like she would do in a printed workbook. All interactions between the learner and the tool are automatically recorded by the SMALA logging service. Not only the data entered by the learner, but also any feedback or hint requests are stored chronologically and time-stamped along with the automatic assessment results and feedback messages that are displayed to the learner. A typical feedback message and the related highlighting of the erroneous step in the ComIn-M user interface are shown in Fig. 8.7.

In this example, the assessment detected that the submitted solution does not utilize the induction hypothesis stated earlier in the proof. The tool informs the user about this finding and offers an additional hint for helping the learner to resolve this issue. However, the learner never receives a bottom-out hint revealing the concrete solution to the problem. If ComIn-M's automatic assessment fails to identify the reason for an erroneous solution step, the tool explicitly recommends reporting the problem to the assigned teacher. Now, it is the choice of the learner whether she tries again and resolves potential errors on her own or whether she agrees to sending the request and, optionally, add some personal questions that remained open at the current stage of the learning process. By simply submitting the *Contact tutor* dialog box from within the ComIn-M exercise sheet (see Fig. 8.8), the teacher gets notified by the SMALA infrastructure and automatically obtains all information that is necessary to reconstruct the learning process of the requesting student (see Sect. 4.3.2). All subsequent direct tutoring between teacher and learner happens asynchronously by email and represents the "human" part of the assessment process incorporated in the SMALA toolset.

Example 2: Squiggle-M—Investigating the concept of functions and relations

Idea. The learning software Squiggle-M provides an interactive learning environment for investigating the concepts of functions, relations, and their characteristics. Using different kinds of virtual learning laboratories, the student can interactively define and manipulate relations, explore different graphical representations of functions, and test her knowledge on functions and their characteristics. Particularly, the assignment laboratory and the representation laboratory invite the

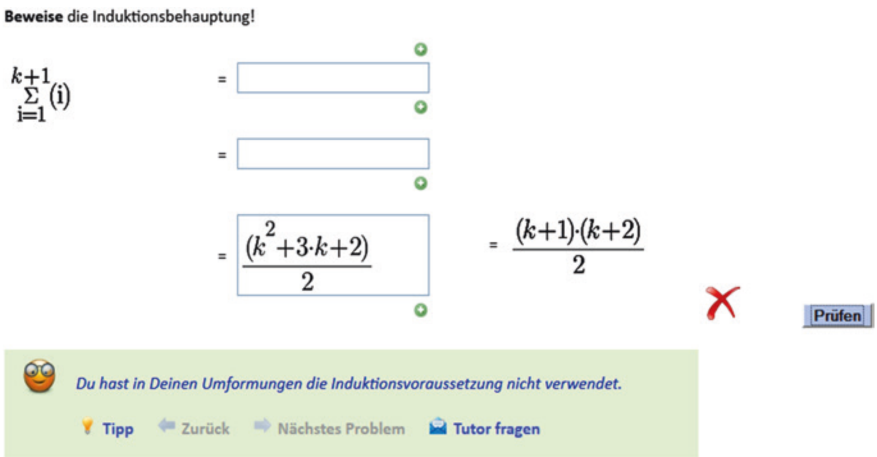

Fig. 8.7 Automatic feedback in the students' view

Fig. 8.8 Contact tutor dialog box

student to examine self-defined functions and relations. Integrated research questions guide the student through the laboratories and can be used as starting points for using the tool. The main focus of the learning tool Squiggle-M is to offer different ways of gaining an extended understanding of the notion of functions, by providing learners with novel—and possibly revealing—visual representations of mathematical concepts and their relation to each other.

Semi-automatic assessment and feedback. Squiggle-M addresses the creativity and curiosity of the learners by letting them define their own assignments and functions. The learners can use the automatic feedback feature of the tool to have these assignments analyzed and obtain information on the assignment's properties. Similarly, the learners can enter their own function equations and have them depicted in different visual representations. As a special feature of the learning tool, the transition from one representation to the other is shown using graphical animations. Assessment in these exploratory-oriented tasks primarily aims at supporting the learners in their discovery process by providing them with extended information on the examined function. Additionally, Squiggle-M offers more concrete problems and questions that have to be solved by the users. Typically, these problems are presented as multiple-choice questions or sequences of them. An example of such a situation is depicted in Fig. 8.1.

After solving a problem, the learner can request immediate feedback on her answer and can optionally retrieve a hint for getting further help. As with all SMALA-integrated learning tools, user interactions with Squiggle-M are recorded transparently by the SMALA logging service. Thereby, it is possible to trace a learner's path through the Squiggle-M laboratories and give individual help or explanations if she gets stuck. Because Squiggle-M relies heavily on visual representations, and because of its technological foundation (a Java applet), steps can be logged with snapshots of the learner's screen. Thus, the teacher has the same view on the assignment or function as the learner has and can analyze the current situation. The capturing of the current visual representation is also offered as so-called “camera” feature in Squiggle-M: By pressing the camera button, the user can save a snapshot of the current assignment for later use.

8.4.3 Analytical Processing

In addition to the students’ view on the learning tools, the SMALA service offers an analytics view on the learning tools usage. Analytical processing occurs in different places in the SMALA system architecture: First, it is realized in the learning tools themselves, and second, it is done in the commonly used logging service (see Fig. 8.6).

8.4.3.1 Analytical Processing in the Learning Tools

Each learning tool has its very own automatic assessment component that analyzes incoming solutions according to domain-specific criteria. Results from this analysis are reported as immediate feedback to the learner. As described in the section above, the feedback is provided as conversational style text and directly addresses the learner (see *Personalization Principle* by Clark and Mayer 2011). For every detected problem, one feedback message and one or more hints are generated. It is up to the learner to decide how many feedback messages or hints are necessary for her to move on in the problem-solving process. Automatic assessment results are not only reported to the learner, but also recorded as assessment events by the SMALA logging service. Assessment events include information about the analyzed solution, detected problems, erroneous steps in the solution, and feedback messages displayed to the user. By doing so, the whole process of a learner’s activities and related tool assessment results and feedback is stored persistently in a central place. This store of event data is the basis for all further analysis that is performed by the analytics components in the SMALA server.

8.4.3.2 Analytical Processing in the SMALA Server

All learning data from the SMALA learning tools are collected in real time by the SMALA logging service. Essentially, the service provides two types of analysis views on the data: the views on the individual learning processes and the views on the overall learning activity and group performance. Individual learning processes are most interesting in the case of personal feedback requests. After the learner has agreed to submit a personal feedback request, the SMALA infrastructure automatically sends a notification email to the responsible teacher. As can be seen in Fig. 8.9, the email contains a link to the SMALA Web service.

Fig. 8.9 The mail received by the teacher indicating the help request

By following this link, the teacher is directly presented with a view of the student's session recording. All user input data and actions as well as automatically generated assessment information are listed in chronological order as a history of interaction events. Figure 8.10 shows an extract from an example recording of a ComIn-M user session.

Using both the information on the learner's line of action and the results from the automatic assessment, the teacher can draw her own conclusions on the individual learning process and give personal advice and feedback to the learner. Even in an asynchronous learning arrangement like this, the teacher can follow the individual steps of the learner and gets additional support from the automatic assessment system that marks and annotates erroneous steps in the process as can be seen in Fig. 8.10.

8.4.3.3 Overall Learning Progress and Group Performance

Obtaining timely feedback on the overall learning progress is essential for teachers to take corrective measures and adapt teaching to the current situation. Especially, in a university setting where courses have a lot of participants, it is very difficult to get information on the group performance early on in the semester. Typically, it is only at the end-of-term examinations where problems and misunderstandings become obvious. In order to address this issue, the learning analytics component of the SMALA infrastructure provides teachers with suitable overview tables and visualizations of the learning events that are generated by the interactive learning tools. Different views provide insights into the data from a high-level overview of tracked learning activities to more detailed views such as detected error types. Linking of related data enables the teacher to drill down from the general views to the more detailed views. For example, it is possible to navigate from the overview of all solved exercises to the session list for one selected user so as to investigate samples and find out why a given error type was common. From the session list, the teacher can further drill down to the chronological display of all events of an individual user session. Moreover, automatic assessment results from the learning tools are aggregated and analyzed in a way that allows the teacher to see, which

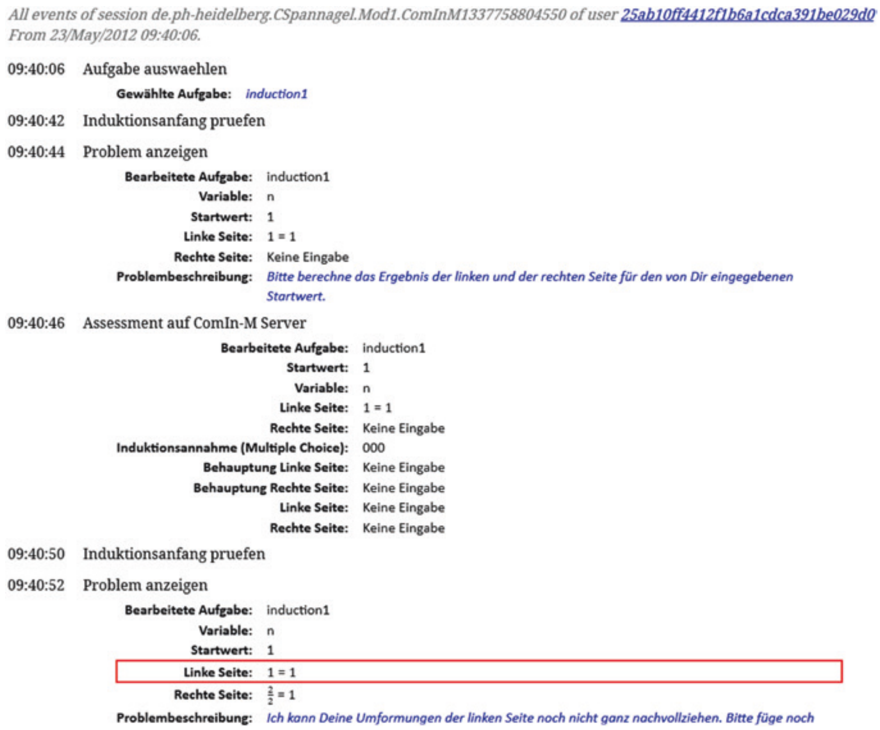

Fig. 8.10 The log view of an individual learning process

error types are most frequent among individual users or how many errors of a certain type were reported for the learning group in total. An example of an aggregated view is depicted in Fig. 8.11.

Based on this information, the teacher can identify common problem areas that might be worth discussing in class again or that might be worth reporting to the learning tool developers. What is more, the teacher can use the log views to monitor impacts of didactical measures in class: How effective are repetition sessions? Are students more successful in solving exercises after a repetition? Can in-class tool demonstrations motivate students to do their own explorations in the learning environment at home? The SMALA analytics view aims at supporting the teachers to answer these kinds of questions and helping them to continuously reflect and improve their own teaching.

8.5 Evaluation and Discussion

In the winter term 2011/2012, the SMALA-integrated learning tools such as ComIn-M, Squiggle-M, and MoveIt-M were used and evaluated in the universities of education of Heidelberg, Karlsruhe, and Ludwigsburg (Germany). In order

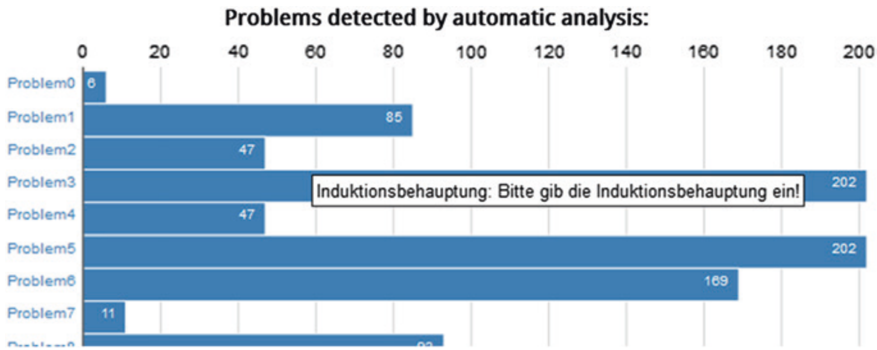

Fig. 8.11 Frequency of error types

to ensure a smooth integration in the established learning platforms of all three universities, deployment scripts and guidelines were developed for StudIP and Moodle. For every participating course, an individual learning activity was set up in the SMALA environment that is only accessible by members of the registered course within one university. By doing so, data security and compliance with strict access control requirements on the learning data could be ensured.

The tools were used in a learning scenario where the tools' usage was first demonstrated in class (by the teacher) and then used by the students for doing their homework. According to the didactical design pattern *Technology on demand* (Bescherer and Spannagel 2009), students were free to choose whether they wanted to use the learning tools to solve their exercises or not. Thus, using the tools was not mandatory in the setup.

Various aspects of implementing the SMALA model in a real-life scenario were considered in the evaluation. First of all, the evaluation should prove whether the SMALA toolset can be integrated in the existing everyday technological infrastructure of the participating universities. Applying the before-mentioned scripts and guidelines, all teachers were able to set up learning activities within their learning management system that offered the SMALA learning tools to their course participants. By doing so, students could log in the learning management system (LMS) and start the tools from within the LMS environment. Pseudonyms were automatically generated for every user and were used as credentials for working with the learning tools. As a result, we can state that the LMS integration for StudIP and Moodle is feasible and could be realized successfully. During the evaluation run of the SMALA toolset, the scalability of the SMALA infrastructure reached boundaries, but these could be resolved. In total, 24,655 events were recorded by the SMALA logging service during the evaluation by 156 users having run 965 sessions.

Another important aspect in the evaluation was the extra workload for teachers due to feedback requests to the teacher. Contrary to what was feared, this did not become a hurdle. We counted a maximum of eight personal feedback requests per tool and course. So the additional workload for analyzing and answering these requests was minimal. According to the teachers, the process recordings and the generated views on these processes were precise enough for them to easily identify mistakes in the solution process and explain the learner how to resolve them. Therefore, we conclude that the granularity of the process recording and the presentation of the solution steps were adequate for reproducing the learner's problem-solving strategy.

After the evaluation period, the learners were asked to give feedback on their experience in using the SMALA-enabled learning tools. The evaluation of the learners' questionnaires showed that many students appreciated the interactive learning tools as additional learning opportunities and that they liked the stepwise assessment and feedback feature of the toolset. However, all students indicated that they preferred pencil and paper for solving mathematical problems to using a computer for mathematical problem solving. In the same way, a vast majority of students prefer asking peers for help or face-to-face tutoring in exercise sessions to using the "Contact tutor" feature offered by the learning tools. In the evaluation group from Karlsruhe, some of the participants criticized missing bottom-out hints with sample solutions. Based on this feedback, we conclude that the majority of students are still reluctant to embrace new technology in learning environments that traditionally used to work with pencil and paper, sample solutions, and exercise lessons in the classroom. Thus, motivating and encouraging students to use new technology is a challenging task for teachers. In order to introduce innovative technology-based learning materials successfully, it is essential that the didactical setup is prepared thoroughly and that special care is taken to make the transition as smooth and easy as possible.

Finally, we collected feedback from the four participating teachers in the form of semi-structured interviews. All teachers agreed that the log views on individual learning processes are very helpful when answering personal feedback requests. However, due to the vast amount of recorded data, none of the teachers has actually tried to get an insight into the overall learning progress by following multiple individual sessions. In order to get an overview of the performance of the whole learning group, the teachers generally agreed that automatically created summaries and suitable visualizations of the data are necessary. At the time of the evaluation, only summary views for the learning tool ComIn-M were available, so the need for more tool-specific summary views became obvious. As a concrete requirement, one interviewee requested an overview of all solved exercises per user (and whether the solutions were correct or not). Ideally, aggregated views on common error types detected by the automatic assessment components of the learning tools should be provided as well. Based on these results, we have proposed various interactive visualizations of the learning data (Rebholz et al. 2013) that are prototypically implemented in the current version of the SMALA infrastructure.

8.6 Conclusion

Letting automatically assessed learning tools be used in and out of classroom learning is an ongoing wave whose effectiveness has been repeatedly evaluated. Few studies, however, have reported more than a *changed teacher involvement*. In particular, the few studies report how teachers can provide feedback to the students' work out of classroom. Our approach describes such a possibility in a way that may bring back the teacher closer to the usage of the learning tools. This possibility transforms the learning tools into smart learning environments which can provide feedback in a relevant and context-specific way, using automated analysis or teacher-lead analysis.

The role of the teacher in the teaching analytics scenarios as we have described includes the following:

- Introduce the usage of the learning tools, in connection with other parts of the courses, such as theory presentations, other learning tools, or expected assignments (pattern *Technology on Demand*, Bescherer and Spannagel (2009)).
- Make sure the learning tool is easily accessible by students by linking to it appropriately in such an environment as the learning management system.
- Encourage the usage of the learning tools in the relevant times of the learning process (e.g., by assignments, by organizing in-room training).
- Employ the analytical views to evaluate the impact of the learning as can be seen in the learning tools' usage:
  - Globally, live in classroom, employing only views that do not show individual actions,
  - Globally, and individually anonymously for sampling, when planning subsequent courses,
  - Individually with known identity, when help is being requested.
- Revisit learning tools' usage and formulate suggestions to developers or adjust instructions to enhance the quality of the learning tools.

In this chapter, we have described the instructional approach underlying the deployment of learning tools in such a way that they can transparently send logs describing the learning process visible in the learning tool. The feedback production is organized in such a way that predictable feedback can be provided automatically after an automatic analysis of the user's input, while feedback that needs more expertise is requested from the teacher.

Experiments that we have run show the technical feasibility of a smart support for teachers and learners. This support is described in the SMALA model for deploying learning tools within a traditional higher education setting so that the learners are neither requested to log in nor requested to agree to a disclosure agreement: Because the learning tools are directly integrated in the learning management system, they only need to be logged in there; because of the storage of the logs uses pseudonyms that cannot be converted to any personalized information, the logs are not considered personal data and are thus not subject to most of the regulations that require, for example, their removal after a short time and the explicit agreement of the learners. Nonetheless, the normal Web page displayed before launching the learning tool indicates to the learners that their usage will be logged, and some learning tools offer the possibility to switch off logging temporarily. Questionnaires distributed after the usage periods of the learning tools have indicated that no significant concerns about the privacy were expressed by the 146 students of the 156.

On the side of the teacher, the implementation of the logging views for the learning tools described in this paper has proven effective and expressive enough. It has been shown that the log views allow teachers to properly understand the individual learning process and effectively provide feedback; the logs being collected, aside of the screenshot at time of sending the request for Squiggle-M, included each of the attempts of the learners and each of the feedbacks. The display of the log has been adjusted for each of the tools so that the actions are almost as expressive as a student's screen. In the case of ComIn-M, the inputs include mathematical expressions, which are then displayed in the log views. Even though the formulae were stored in OpenMath and the display was made in MathML, one of the teachers asserted that the *screenshots* of the learner's inputs were quite helpful.

8.6.1 Open Questions

Rareness of requested feedback: The offer to formulate questions to the teachers has been considered with a fear of becoming overwhelmed, but no flood happened as very few requests arrived, in comparison to the amount of learning tools' usage or of students. Several hypotheses can be formulated to explain this fact: The first is the preference to ask in presence of their peers and teachers or tutors (and indeed, this has been the majority of answers to this question), and the second is the possibility that simply writing the question helps the learners finding the answers and thus stops their question writing process. A finer grained analysis would be needed to elucidate the best strategy to motivate asking questions productively.

Didactical Usage Patterns: The set of didactical configurations where the analytics views can be employed is not entirely investigated:

- Clearly, the personal log can be useful for learners themselves to support a reflection on their own learning process and has been made available, but no teacher encouragement to take such actions has been made, and thus, it has almost not been used.
- Peers and other persons in the learners' circles may take advantage of such log views too: Within an electronic communication among peers so as to exchange discoveries or solution paths, in family circles so as to discuss and help one's own child's learning. The conditions and best practices of doing so similarly to the approach of ePortfolios' assembly of receipts that prove the learning (Ravet 2009) could be researched.
- Classrooms may take advantage of some of the logging views. A teacher may want to show a live view while the exercises are being run, so as to show the class's progress. Is there a risk of the anonymity breach? Should teachers be endowed with a special mode so as to avoid inadvertently show individual data?

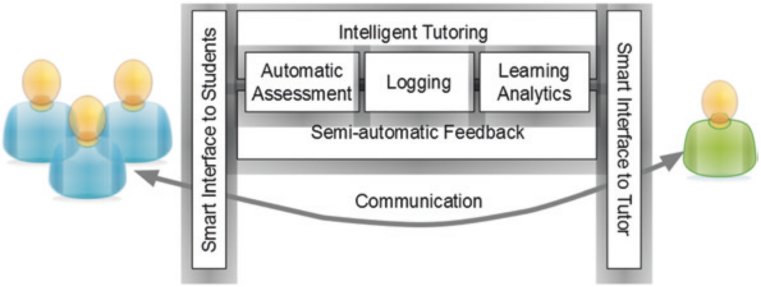

Fig. 8.12 Smart interaction—Interplay between learners, teachers, and technology

Such a view is clearly helpful when formulating an encouragement to the class to use the learning tool; is it different if displayed live in the classroom when the tools are being used?

Quantity of Logging Information: The amount of logging events, and the information inside each log event, is another uncertain variable. A complete video of the learning tools' usage is clearly too detailed to get a quick overview and probably too heavy to be processed quickly. But simple scores stating the results are clearly too light to provide an explanation. The approach we described stands in the middle between these two extremes. It needs to be sufficiently rarely sent so that a usage session holds in a few screen-pages, but it needs to contain sufficient information so that one can conclude what the user has performed as action. In such a system as ComIn-M, the learner's input formula is an effective representation, but representing as a formula such an input as the relationship between elements of two sets in Squiggle-M is probably too compact to be effectively read. Thus far, our only criteria for the informativeness of a logging event's display have been that it resembles the user's input or view. What are other criteria for other learning tools? Is the log display of the car simulator described in TinCanAPI's story[3] sufficiently effective to show the simulator usage to a teacher that knows the simulator a bit? How could it be done for a dynamic geometry system?

Privacy: Furthermore, the best practice to ensure a feeling of privacy among the students' remains to be defined. For example, we have observed that some tutors started to progressively remember the pseudonym of learners and to formulate expectations when drilling down from global views to individual views. This contradicts the role of pseudonyms, which, precisely, are meant to hide the identity of the learners. It may be that more anonymization is required in the log display (so that remembering cannot be done) and in the URL of the log views, or it may be that ethical guidelines should be expressed to teachers so that bad surprises such as the mention of a typical error in an observed session within the classroom course can not occur.

[3] The log display is a large table expressing the events with details; see http://tincanapi.com/a-simulators-story/.

8.6.2 Vision

To sum up, we present our vision of a smart learning environment: an environment that combines intelligent tutoring enhanced by human expertise and learning analytics enhanced by human analytical skills in one system.

As Fig. 8.12 shows, smart interfaces wrap the diversity of learning tool-specific assessment, logging, and learning analytics components and provide a homogeneous view on the learning content and the recorded learning data, respectively. Ideally, this environment supports the interplay between learners, teachers, and technology in such a way that smart interaction becomes reality.

References

Ainsworth, L., & Viegut, D. (2006). *Common formative assessments*. Thousand Oaks, CA: Corwin Press.

Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence, 42*, 7–49.

Anderson, J. R., & Pelletier, R. (1991). A developmental system for model-tracing tutors. In *The International Conference on the Learning Sciences* (pp. 1–8), Charlottesville, Virginia.

Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly, 33*(1). http://www.educause.edu/ero/article/signals-applying-academic-analytics

Bannan-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher, 32*(1), 21–24.

Bescherer, C., Kortenkamp, U., Müller, W., & Spannagel, C. (2009). Intelligent computer-aided assessment in math classrooms. In A. E. A. McDougall (Ed.), *Researching IT in education: Theory, practice and future directions* (pp. 200–205). Milton Park, New York: Routledge.

Bescherer, C., & Spannagel, C. (2009). Design patterns for the use of technology in introductory mathematics tutorials. In A. Tatnall & A. Jones (Eds.), *Education and technology for a better world* (pp. 427–435). Berlin: Springer.

Bescherer, C., Herding, D., Kortenkamp, U., Müller, W., & Zimmermann, M. (2011). e-learning tools with intelligent assessment and feedback. In K. Sabine Graf, F. Lin, & R. McGreal (Eds.), *e-learning tools with intelligent assessment and feedback* (pp. 151–163). Hershey: IGI Global.

Brown, S. (2004). Assessment for learning. *Learning and Teaching in Higher Education 1*, 81–89.

Buchem, I., & Pérez-Sanagustín, M. (2013). Personal learning environments in smart cities: Current approaches and future scenarios. *eLearning Papers, 35*. http://openeducationeuropa.eu/sites/default/files/asset/In-depth_35_1.pdf

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. R. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*(4), 273–281. http://learninglab.psych.purdue.edu/downloads/2007_Butler_Karpicke_Roediger_JEPA.pdf

Carnegie Mellon University (2013). CTAT—Cognitive Tutor Authoring Tools. Retrieved January 3rd, 2014 from http://ctat.pact.cs.cmu.edu

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (3rd ed.). San Francisco, CA: Wiley.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997a). Chapter 37: Intelligent tutoring systems. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human–computer interaction* (pp. 849–874). Elsevier Science B. V, Amsterdam.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997b). Intelligent tutoring systems. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction*. Amsterdam: Elsevier Science B. V.

Dodds, P., & Fletcher, J. D. (2004). Opportunities for new "smart" learning environments enabled by next generation web capabilities. *Journal of Educational Multimedia and Hypermedia, 13*(4), 391–404.

Essa, A., & Ayad, H. (2012). Improving student success using predictive models and data visualisations. *Research in Learning Technology, 20*(0). http://www.researchinlearningtechnology.net/index.php/rlt/article/view/19191

Feng, M., Heffernan, N. T., & Koedinger, K. R. (2009). Addressing the assessment challenge in an Intelligent Tutoring System that tutors as it assesses. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI Journal), 19*(3), 243–266. http://web.cs.wpi.edu/~mfeng/pub/USER562.pdf

Guéraud, V., & Cagnat, J.-M. (2006). Automatic semantic activity monitoring of distance learners guided by pedagogical scenarios. In *Proceedings of EC-TEL* (pp. 476–481). Springer, Berlin.

Hattie, J. (2008). *Visible learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 81–112*(1). http://growthmindseteaz.org/files/Power_of_Feedback_JHattie.pdf

Heffernan, N., Heffernan, C., Dietz, K., Soffer, D., Pellegrino, J. W., & Goldman, S. R. (2012). Improving mathematical learning outcomes through automatic reassessment and relearning. In *AERA Annual Meeting 2012*.

Jovanovic, J., Devedzic, V., Gasevic, D., Hatala, M., Eap, T., Richards, G., et al. (2007). Using semantic web technologies to analyze learning content. *IEEE Internet Computing, 11*(5), 45–53.

Jovanovic, J., Gasevic, D., Brooks, C., Devedzic, V., Hatala, M., Eap, T., et al. (2008). LOCO-analyst: Semantic web technologies in learning content usage analysis. *International Journal of Continuing Engineering Education and Life Long Learning, 18*(1), 54–76.

Koedinger, K. R., Aleven, V. A. W. M. M., & Heffernan, N. T. (2003). Toward a rapid development environment for cognitive tutors. In *Proceedings of the 11th International Conference on Artificial Intelligence in Education*, AI-ED 2003, Amsterdam, 2003 (pp. 455–457). IOS Press.

Koedinger, K. R., Aleven, V., Heffernan, T., McLaren, B., & Hockenberry, M. (2004). Opening the door to non-programmers: authoring intelligent tutor behavior by demonstration. In *Proceedings of 7th Annual Intelligent Tutoring Systems Conference*, Maceio, Brazil, 2004.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of Feedback and Verbal Learning. *Review of Educational Research, 58*, 79–97.

Mikulecký, P. (2012). Smart environments for smart learning. In *Proceedings of DIVAI 2012—9th International Scientific Conference on Distance Learning in Applied Informatics* (pp. 213–222).

Müller, W., Bescherer, C., Kortenkamp, U., & Spannagel, C. (2006). Intelligent computer-aided assessment in math classrooms: State-of-the-art and perspectives. In *Proceedings of the IFIP WG 3.1, 3.3, and 3.5 Joint Conference Imagining the future for ICT and Education*, Alesund, Norway, 2006.

NEXT-TELL (2013). Call For Papers 3rd International Workshop on Teaching Analytics (IWTA) at E-TEL 2013. Retrieved from http://www.next-tell.eu/teachinganalytics/iwta-2013b/

Ravet, S. (2009). Building my ePortfolio, an ePortolano publication. Available from http://www.itineraires.fr/ePortolano/. Accessed on 15 May 2014.

Rebholz, S., Libbrecht, P., & Müller, W. (2012). Learning analytics as an investigation tool for teaching practicioners. In *Proceedings of the Workshop Towards Theory and Practice of Teaching Analytics 2012 (TaPTA-2012)*, Saarbrücken, Germany, 2012. CEUR-WS.

Rebholz, S., Libbrecht, P., & Müller, W. (2013). Interactive visual exploration of learning data: the role of teachers as learning analysts. In *Proceedings of 10th IFIP World Conference on Computers in Education (WCCE 2013)*, Torun, Poland, 2013.

Rebholz, S., & Zimmermann, M. (2013). Applying computer-aided intelligent assessment in the context of mathematical induction. In Z. Pan, A.-D. Cheok, W. Müller, I. Iurgel, P. Petta, & B. Urban (eds), *Transactions on Edutainment X* (pp. 191–201), Springer, Berlin, LNCS 7775. http://dx.doi.org/10.1007/978-3-642-37919-2_11

Reeves, T. C., Herrington, J., & Oliver, R. (2005). Design research: A socially responsible approach to instructional technology research in higher education. *Journal of Computing in Higher Education, 16*(2), 97–116.

Reusser, K. (1993). Tutoring systems and pedagogical theory: Representational tools for understanding, planning, and reflection in problem solving. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 143–177). London: Routledge.

Siemens, G. (2011). What are learning analytics? Retrieved from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Singh, R., Saleem, M., Pradhan, P., Heffernan, C., Heffernan, N. T., Razzaq, L., Dailey, M. D., O'Connor, C., & Mulcahy, C. (2011). Feedback during web-based homework: The role of hints. In *Artificial Intelligence in Education—15th International Conference, AIED 2011* (pp. 328–336). Springe, Berlin/Heidelberg.

Sleeman, D., B. J. S. (1982). Introduction: Intelligent tutoring systems. In D. S. J. S. Brown (Ed.), *Intelligent tutoring systems* (pp. 1–11). New York: Academic Press.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1),153–189. http://myweb.fsu.edu/vshute/

Tabach, M. (2013). Developing a general framework for instrumental orchestration. In *Conference on European Research on Mathematics Education (CERME8)*.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. http://www.public.asu.edu/~kvanlehn/Stringent/PDF/EffectivenessOfTutoring_Vanlehn.pdf

Vatrapu, R., Teplovs, C., Fujita, N., & Bull, S. (2011). Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 93–98), New York, NY, USA, 2011. ACM.

Worcester Polytechnic Institute (2013). Assistments. Retrieved January, 3rd, 2014 from http://www.aboutus.assistments.org/teacher-formative-assessment.php

Yesner, R. (2012). *The next-generation classroom: smart, interactive and connected learning environments*. IDC Government Insights.

Zhang, H. & Almeroth, K. (2010). Moodog: tracking student activity in online course management systems. *Journal of Interactive Learning Research*, *21*(3), 407–429.

Chapter 9
Interactive Physical Games: Improving Balance in Older Adults

Wei-Chieh Fang, Feng-Ru Sheu, Yu-Lun Lin, Yun-Lin Lee and Nian-Shing Chen

Abstract As part of an ongoing project, we developed a motion-based Evergreen Fitness System (EFS) intending to train balance in older adults. A participatory design approach was applied where health experts were consulted for designing exercises appropriate for balance training, and thirteen older adults (average 68 years old) were invited to use the system for providing feedback. Aspects being evaluated included suitability of the exercises that were integrated into EFS, their enjoyment during the use of EFS and the user experience of the gameplay, system operation, game design, and intent to use. The questionnaire interview shows that our senior participants found the exercises suitable and beneficial and the interactive physical games enjoyable. We also observed that our participants have varying capabilities in performing body motions. They paid more attention on the exercise itself during gameplay than the game results/progress report but would discuss their performance while watching others play afterward. Based on these findings, we proposed that an adaptive mechanism, socially interactive functions and more informative progress report should be included in the system, in hope to promote the optimal physical outcomes.

Keywords Interactive physical games · Exergames · Game design · Kinect · Elderly · Balance exercise

9.1 Introduction

Regular exercise is important for maintaining elder adults' healthy life. Lack of physical activity might lead to higher chances of falls and the development of chronic diseases, such as obesity, high blood pressure, and diabetes (Nelson

W.-C. Fang · F.-R. Sheu · Y.-L. Lin · Y.-L. Lee · N.-S. Chen (✉)
Department of Information Management,
National Sun Yat-Sen University, Kaohsiung, Taiwan
e-mail: nschen@mis.nsysu.edu.tw

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_9

et al. 2007). The danger of falls at home is of particular interest to our research team. As elderly adults' age advances, the decline in muscle power and balance would put elderly adults at higher risk of falling. Once they become fallers, their ability to live alone and maintain functional activities of daily living would be gradually deprived. Moreover, the experience of falls can have psychological effect on an individual's confidence in mobility, leading to activity restriction (Sherrington et al. 2004). Thus, intervention is needed to prevent the incidents of falls.

Several guidelines have been proposed for the prevention of falls. First, regular and persistent exercise has been recommended and has many proven benefits in preventing fall, such as reduced risk of falls and injuries from falls (Nelson et al. 2007). Second, the use of exercise and assistive device collectively as guided by the health professionals can improve balance (Sherrington et al. 2004). Third, exercise directly targeting balance is effective in preventing fall (Nelson et al. 2007).

Despite the recommended exercise for elderly adults, there are some reasons that make it difficult for older adults to engage in physical activities. One is their lack of motivation to exercise (Dishman 1994). Without the sense of social interaction, exercising can be lonely and hard to maintain as a routine. Also, the lack of convenient devices or facilities can make exercise seem troublesome (Schutzer and Graves 2004). Finally, older adults are not offered appropriate exercise guidelines (Schutzer and Graves 2004). As a result, how to exercise appropriately can be a challenge.

Thanks to the advancement of motion-sensing technology, gesture-based games can be used to reduce the aforementioned difficulties. Commercial exergaming technologies, i.e., Nintendo Wii Fit, XBOX Kinect, and Dance Dance Revolutions, have been available, making exercise affordable, entertaining, and socially interactive (Chao et al. 2013; Sheu & Chen 2014). Thus, researchers have started to explore the feasibility of using exergames in promoting physical activity and found that they have positive effects on game results (Williams et al. 2010) and user experience (Gerling et al. 2012), such as motivation (Aarhus et al. 2011), mood (Kahlbaugh et al. 2011), loneliness (Kahlbaugh et al. 2011), and enjoyment (Lange et al. 2009).

However, there is little research that investigated the effectiveness of exergames on the physical fitness outcomes. Besides, these commercial exergames were not designed for the elderly, which might result in undesirable consequences or poor outcomes (Marin et al. 2011). Marin et al. (2011) had commercial games assessed by the health professionals and found that since the older adults' characteristics and needs were not considered during the design process, inappropriate movements might be learned in response to the gameplay, which might lead to undesirable outcomes. This finding suggests that commercial exergames might play an entertaining role rather than a training role in promoting the optimal physical fitness.

Thus, the objective of the current study is to develop a motion-based game system, called Evergreen Fitness System (EFS), combining exercises specially designed for older adults. Specifically, the system aims to improve lower limb strength so as to improve balance. We invited health professionals and elderly participants to jointly participate in the initial design of the system. In order to create exercises suitable for improving older adults' muscle strength of lower limbs, two field experts were consulted. Six exercises were then selected and embedded in the EFS. Then, to explore the feasibility of the system, a questionnaire interview with elderly participants was administered on (1) the suitability of exercises embedded in the system and (2) the user experience of the overall motion-based game system. Finally, the findings from the interview are discussed and will be used to modify the system for its future implementation in a 10-week-long intervention.

9.2 Related Work

9.2.1 Full-Body Interaction Games for Seniors

Unlike traditional videogames using mouse/keyboard, full-body interaction games allow user to control the game using the body through 3D camera, based on which terms such as embodied gaming (Aarhus et al. 2011; Kahlbaugh et al. 2011) and exergaming (Gerling et al. 2012; Billis et al. 2011) have been widely adopted. The affordable mass-produced consoles immediately draw researcher's attention to find out whether they can benefit the well-being and physical fitness in the elderly. The overall findings indicate that older people are receptive to the use of these commercial gaming systems (Ijsselsteijn et al. 2007).

However, there is little research that examined the effectiveness of using exergames on the fitness of older adult (see Table 9.1). Maillot et al. (2012) explored the effects of 12-week Wii Fit on physical functions and found improved performance in healthy older adults (aged 65+) on the functional fitness battery except the Back-Scratch measures of flexibility. The functional fitness battery measures the physiological capacity to perform daily activities safely and independently without undue fatigue. Aarhus et al. (2011) implemented Nintendo Wii Fit in 6-month-long physical training sessions for three groups of seniors (aged 61+) who live independently, who are with physical impairments and who live in the center's nursing home, respectively. Qualitative results on Senior Fitness tests showed that there was a trend in physical improvement of all participating seniors.

Table 9.1 The effectiveness of using exergames

Study	Aim	Game	Measure	Effectiveness
Maillot et al. (2012)	Promote fitness	Wii fit	Functional fitness battery	Improved on most aspects of functional fitness battery except flexibility
Aarhu et al. (2011)	Promote fitness	Wii fit	Senior fitness tests	Qualitative results show improvement on Senior Fitness tests
Williams et al. (2010)	Improve balance	Wii fit	Berg balance scale Tinetti balance assessment tool	Improvement only on Berg balance scale

The senior fitness tests measure the physical capacity needed to perform daily activity for seniors aged sixty and above.

Another study directly targeting the community-dwelling faller (aged 70+) by Williams et al. (2010) compared the effectiveness of attending the Wii Fit exercise sessions with that of attending the local falls groups for 12 weeks. Two balance measures, the Berg Balance Scale and the Tinetti Balance Assessment Tool, were used to assess balance improvements. The results showed that the Wii Fit group (Mean = 48.1) outperformed the local falls group (Mean = 40.1) on the Berg Balance Scale at 4 weeks. Despite the improvement at 4 weeks, the performance decreased significantly for the Wii Fit group at 12 weeks, while there was no significant decrease for the local falls groups, suggesting a lack of sustained improvement for the Wii Fit group.

There are other studies that have developed their own systems for improving lower limb functions in the elderly (see Marin et al. (2011) for a review). Most studies did not adopt the prescribed balance exercise except Doyle et al. (2010) who developed an interactive system that integrated a set of Otago exercises specifically designed for building strength and balance in older adults. It used Shimmer motion sensors to recognize body motions, allowing for home-based exercise to be monitored by the instructors from the exercise program. Flash software was used to provide instructional content, feedback, and navigational tools. It was observed that participants felt more motivated because they knew the whole program was under physiotherapist's guidance. However, some reported that following the program is boring, suggesting that game elements play a role in maintaining continuous exercise.

Based on the findings above, it appears that using exergames is promising in promoting physical fitness. Whether this effect can be extended to the performance of balance in older adults is not yet clear. As our objective is to improve balance, we speculate that the physical exercise involved in the exergames might not sufficiently address this purpose. Furthermore, Marin et al. (2011) found that commercial products would lead to undesired physical conditions for seniors who might perform inappropriate movements in response to the gameplay. Along with the recommendations noted in the exercise literature that "balance exercise" can help reduce falls and that appropriate exercise guidelines should be given, this study intends to design exercises specifically for training balance.

9.2.2 Game Design

Studies have suggested that digital games have great potential for improving mental and physical well-being of older adults. However, most of the digital game design available does not serve older adults well (Ijsselsteijn et al. 2007). What motivates the young-aged group from the game might not work the same way for seniors. Using player-centered design research, Schutzer and Vandenabeele (2008) found that older adults consider meaningful play more important than mere entertainment arising from the games. Among the important game concepts provided by seniors, purpose in the game was greatly valued. The older adults showed great interests in receiving educational or cultural benefits, such as gaining knowledge or doing exercise for the purpose of maintaining health (Schutter and Vandenabeele 2008).

Similarly, Ijsselsteijn et al. (2007) argued that perceived benefits are important when it comes to older adults' willingness to engage in the digital game. A study comparing several Nintendo Wii games found that older adults aged between 62 and 89 enjoyed the games in which they can see their progress as a result of better body control, more than other types (Aarhus et al. 2011). Seniors enjoyed the games more as they improved their skills in the games and advanced to the next level using the body. Thus, it is suggested that when designing a digital game for seniors, focus should be put on "progress and mastery rather than ability judgments, which will likely increase the senior's motivation and persistence to engage in a task…" (Ijsselsteijn et al. 2007). In other words, game elements that encourage participation work better than the common game element and competition for seniors.

Another issue that might affect the use of digital game is usability. Due to the fact that there are age-related declines in sensory–perceptual processes, cognitive processes, motor abilities and response speed, digital games designed for the young-aged group are not suitable for the seniors (Ijsselsteijn et al. 2007). For example, seniors might find it hard to follow the dynamic visual animation, pace of the gameplay, or even the range of motion required by the game.

Furthermore, seniors might lack the mental representation of how computer technology works (Ijsselsteijn et al. 2007). Gerling et al. (2012) compared older adults' usability experience of Wii Fit with their own-designed SilverBalance which features gradual increase in game speed and easy entry into play without much gaming experience. It was found that SilverBalance provided better game control and better support during exercise. They also observed that additional feedback was needed to facilitate interaction with the game and that graphical information on screen should be reduced to allow for optimal concentration on body movements.

These findings suggest that older adults have preferences over certain game elements, such as meaningful content, expect to gain some benefits from using the new technology and have special needs for easy operation. Based on the literature above, this study designed a game that is meaningful (e.g., the provision of travel theme), beneficial (e.g., teaching them the correct movements for improving balance and showing progress), and elder-friendly.

9.3 Design and Prototype

The aim of this study was to develop a motion-based system to strengthen senior's balance performance. Specifically, the system integrated exercises prescribed by the field experts and featured physical practice through interactive games. Since the target users are the older adults who are over 60 years old with special needs to be addressed as aforementioned, the interactive system was carefully designed. This study first consulted two field experts to derive proper exercises for improving balance performance, then developed the EFS, a home-based interactive prototype, and finally invited senior participants to jointly evaluate the suitability of the movements and the enjoyment of using the system, as well as provide feedback to improve the prototype.

9.3.1 Exercises for Older Adults

As the goal is to improve senior's balance performance, the focus of these exercises is on lower body strength. Six exercises were derived according to field experts' suggestions as follows (Table 9.2). These exercises are the ones that are commonly instructed to self-perform at home for the prevention of falls among older adults, which are adopted as the core exercises in the system.

9.3.2 Evergreen Fitness System (EFS)

9.3.2.1 System Architecture

EFS is a game-based prototype developed to train older adults' balance using the Microsoft Kinect sensor. The Kinect sensor is a structured-light 3D senor that can recognize body gestures and body motions. Gestures are used to select the menu, while body motions are required as part of the exercise. The system software was designed using Microsoft Kinect SDK 1.6 (Software Development Kit). The system features physical interaction and a travel-themed game that prompts a series of the aforementioned strength exercises purposefully designed for the elderly. Immediate feedback, such as smiling face, ding sound, and entertaining background music, is provided to encourage interaction.

Table 9.2 Six exercises designed for improving balance

Number	Exercises	Function
Exercise 1	Knee marching	Strengthen ankles and hips; improve dynamic or moving balance
Exercise 2	Side hip raise	Strengthen side hip muscles; maintain lower body endurance to better walk and side step around objects
Exercise 3	Lunges	Strengthen quadriceps and hips; improve the ability to get out of a chair and balance
Exercise 4	Partial squats	Increase hip flexibility, quadriceps strength and hip flexor strength; steady body for better balance and safety
Exercise 5	Wide squats	
Exercise 6	Standing knee flexion	Strengthen hamstring muscles; help with standing balance

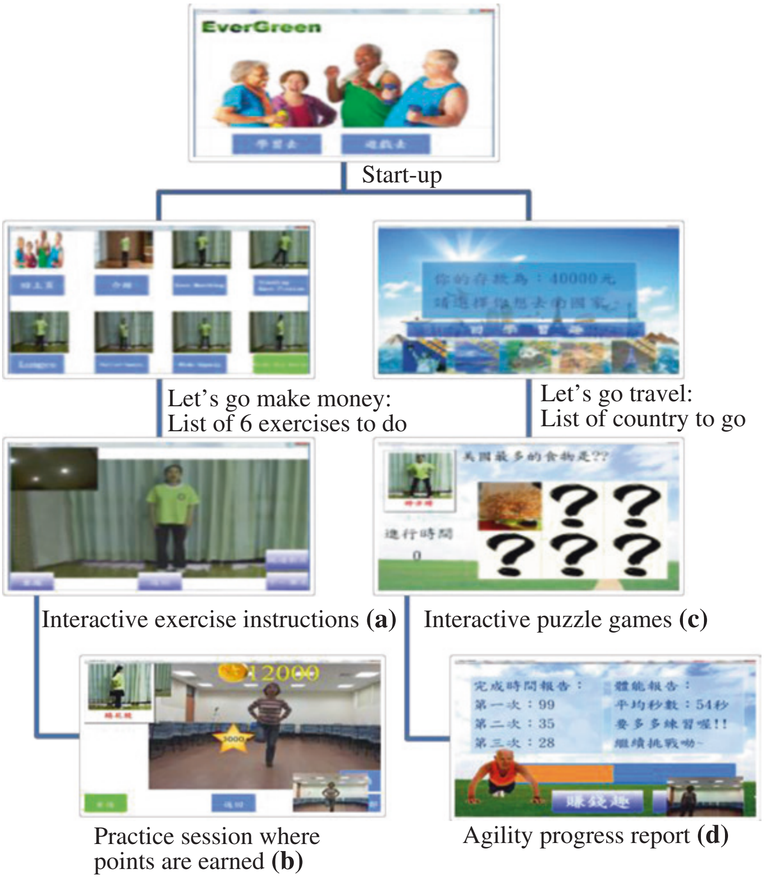

Fig. 9.1 System content of evergreen

9.3.2.2 System Content

The prototype includes two phases (See Fig. 9.1): The first phase was designed to deliver interactive video-based exercises for training balance, and the second phase was designed to elicit optimal exercise through gameplay. The two phases were integrated into a travel-themed game. In the game mode, the first phase is presented as *Let's go make money* and the second phase as *Let's go travel*.

9.3.2.3 Let's Go Make Money Phase

In this phase, the system will assist senior users to learn the six core exercises designed to enhance balance. Brief exercise knowledge about balance performance and the six core exercises will be presented in the form of embedded videos. Older adults are allowed to select and begin with the most comfortable exercise. While watching the exercise video, they can see themselves and follow it (Fig. 9.1a). The system will give visual and audio feedback. Immediately following the exercise instruction, they will be required to perform the exercise on their won (Fig. 9.1b).

To motivate engagement, virtual money can be earned upon the completion of a correct movement performed. The virtual money accumulated can be used for the subsequent gameplay. The purpose of this phase is to instruct senior players on the core exercises.

9.3.2.4 Let's Go Travel Phase

In this phase, senior players will be required of practicing the exercises through a series of travel-themed puzzle games (Fig. 9.1c). Using the virtual money earned from previous phase, they can choose a foreign country to tour. Each tour is composed of three interactive puzzle games. Each game has a hidden picture, along with cultural questions, such as "What is the most common food in the US?" The picture will be gradually revealed, such as the picture of burger, when the exercises prompted by the system are correctly performed.

All the interactive puzzle games will give insights into the food, culture, and attractions of the county chosen. Audio feedback, ding sound, will be provided to reinforce the completion of the exercises. The time spent on solving the three puzzle games will be recorded, and each score will be presented respectively as an agility progress report at the end of the game (Fig. 9.1d). The objective in this phase is to motivate the elderly to exercise with meaningful play—gaining cultural knowledge and perceiving improvement.

9.4 Evaluation and Procedure

The goal of the evaluation is to include the target audience during the design process. To better design the EFS, this study adopted a participatory design (Fig. 9.2). Since seniors have their limitations in performing a certain range of motions, it is crucial that we do not set the predefined motions too physically challenging. Thus, we examined the suitability of exercises embedded in the system. Furthermore, to understand whether senior users' experience of using Evergreen was positive, we examined the user experience and measured their exercise enjoyment.

Thirteen seniors were recruited from an Evergreen College in Southern Taiwan, and their characteristics were collected. During the participatory design session, a Physical Activity Readiness Questionnaire was first administered to ensure that they were physically and mentally ready for the following activities. Each was then invited to use EFS, while others were watching and following as they liked (Figs. 9.3 and 9.4). Following that, their exercise enjoyment throughout the use of the EFS was measured using the Physical Activity Enjoyment Scale (PAES) (Raedeke 2007). Finally, a semi-structured interview was conducted to examine

Fig. 9.2 Participatory design

Fig. 9.3 Participant exercising in the game session

Fig. 9.4 Participants watching while others playing the game

the suitability of these 6 embedded exercises and explore user experience of EFS. The whole session took about three hours with each participant[1] spending around fifteen minutes on the system.

[1] Video showing participants using Evergreen Fitness System: http://www.youtube.com/watch?v=PJTUShQn3eg&feature=youtu.be

9.5 Results

9.5.1 Evergreen Fitness System (EFS)

Thirteen seniors participated in the study with an average age of 68 (SD = 6.81, range: 60–80), 2 males and 11 females (see Table 9.3 for description of the participants). In terms of exercise habit, they are active in exercise, who exercise 40 min per time and 4 days a week on average. Seventy-six percentage of them use no assistive exercise device, 62 % prefer aerobic exercise, 62 % exercise alone, and 50 % exercise for maintaining health. In terms of computer use experience, they have average 5.58 years of experience in computer (SD = 11.1, range: 0–40) and 0.92 year in video games (SD = 1.73, range: 0–6), suggesting that they have little gaming experience. Only one of all the participants had experience in playing exergames. The average year of education is 10.69.

9.5.2 Physical Activity Enjoyment Scale (PAES)

PAES, a validated and reliable measure of exercise enjoyment ($r = 0.94$) (Raedeke 2007), was used to understand how much our target audience enjoyed the exercise during the gameplay. There were 8 items reflecting generalized state of enjoying the activity and experience they underwent (i.e., I enjoyed it; I felt interested; I liked it; I found it pleasurable; It was a lot of fun; It was very pleasant; I felt as though there was nothing else I'd rather be doing; I was very absorbed in the activity). On a 7-point Likert scale, with scores higher than 4 representing greater enjoyment (from 1 = not at all to 7 = very much), the average enjoyment scores were 6.35 (SD = 0.92), suggesting that the senior participants enjoyed and engaged greatly in using the system for exercises.

Table 9.3 Sample characteristics

Variable	Mean (SD)
Age	68 (SD = 6.81)
Gender	2 males
	11 females
Exercise frequency	4 days a week
	40 min each time
Computer use experience	5.58 years (SD = 11.1)
Video game experience	0.92 years (SD = 1.73)
Education	4 college education
	4 high school graduate
	2 junior high school graduate
	3 elementary school graduate

9.5.3 Suitability of the 6 Embedded Exercises

A questionnaire interview was used to evaluate whether the senior participants were capable of performing the 6 exercises required by the system. They were asked to rate their interaction experience on a 5-point Likert scale. Table 9.4 gives an overview of descriptive results of the questionnaire.

Results show that on average their responses to the questions were positive as the ratings on all the questions were above 3. This suggests that they were capable of performing the exercises required by the system, as they found these exercises were doable ($M = 4.5$), safe ($M = 4$), easy ($M = 4.58$), and not tiresome ($M = 4.58$). However, in terms of movement detection, the rating was relatively low ($M = 3.75$), suggesting that their movements were not perfectly recognized by the system.

Feedback from the interview shows that while 58 % of them perceived the exercises easy to perform, 42 % found certain movements hard to perform, and some of whom were even afraid of falling. These movements were as follows: standing knee flexion, partial squats, and side hip raise. However, some reported that it is not that they were not able to perform these exercises, but that their limited range of motions did not trigger the system. In response to this issue, some suggested that (1) motion detection should not be too rigid and that (2) the system should provide “corrective feedback,” such as showing a visual prompt like lift your leg a bit higher, to guide them in adjusting their movements instead of leaving them to try out on their own until it is detected.

Most senior participants reported that the interactive feature and real-time feedback, such as providing auditory and visual feedback, was very useful in encouraging the completion of exercises. Overall, many reported that they were capable of completing most of the exercises and would need some practice for certain movements that appeared challenging to them.

9.5.4 User Experience of EFS

To explore senior users’ experience of using EFS to exercise, they were asked to rate: system operation, game design, and intent to use the system. Results showed

Table 9.4 Descriptive results for evaluating the suitability of the exercises embedded in EFS

Item	Mean	SD
Q1. The exercises were doable	4.50	0.52
Q2. I was not afraid of injury when practicing the exercises	4.00	1.13
Q3. Practicing the exercises was not tiresome	4.58	0.52
Q4. Practicing the exercises was easy	4.58	0.52
Q5. My movements were correctly detected	3.75	1.22

Table 9.5 Descriptive results for evaluating user experience of EFS

Item	Mean	SD
Q1. The system operation throughout gameplay was easy	4.08	0.90
Q2. Practicing the exercises through gameplay was not tiresome	4.50	0.52
Q3. Practicing through gameplay was fun	4.58	0.90
Q4. Practicing the exercises through gameplay was challenging	4.50	0.52
Q5. Practicing the exercises through gameplay was motivating	4.67	0.49
Q6. Practicing the exercises with the system is beneficial to my balance	4.58	0.52
Q7. Assuming I have access to the system, I intend to use it	4.58	0.79

that their experience of using EFS was highly positive as the ratings on all the questions were above 4, see Table 9.5.

Despite the high ratings, feedback from the interview and our observations provide deeper understanding for improving the system as noted below.

9.5.4.1 System Operation

Our senior participants appeared to find the motion-sensing interface convenient. They were delighted that they did not have to wear any devices during the exercise, and the feature that they could see themselves interacting with the system on the screen helped them to better perform the movements required. However, in terms of menu selection, it appeared that selecting the menu by waving hands to the sides to shift between icons and by clapping hands to select the icons was not easy for them. Since we were informed that they did not have adequate experience in using gesture-based technology, we designed this conventional way of selecting menu for it approximates the graphical user interfaces on the desktop. It turned out that it required additional learning and practice.

We believe that as they become more experienced in using their body to command, this will not be an issue. Also, it was found that a timely prompt would be helpful when the senior participants are idle or do not follow the system. Some noted, “we were not sure about what to do next,” and “we don’t know what to do to control the system.”

9.5.4.2 Game Design

Some participants reported that they liked the travel theme. They enjoyed seeing scenery around the world and reading the intriguing descriptions while playing games. Some reported that “it would be great if more beautiful pictures can be added.” Some reported, “I’d like the content to be more diversified” and “it would be better if the theme can be more relevant to my life, such as gardening and learning to write calligraphy.” It appears that providing a variety of themes would make the game more appealing. Themes reported to be interesting were art craft making, painting, and cooking.

Moreover, most reported that the game was engaging. However, rather than paying attention to the scoring system, such as virtual money earned and the agility progress report, they were focused on the completion of the exercise itself. Some even reported that they were not aware of the timer shown on the screen during gameplay. Many suggested that it would be more challenging if the game provides different levels or complexity of exercises to choose from. It appears that they were concerned about the exercise tasks instead of the results of the game performance. However, we observed that those who watched others playing would talk about their game performance and to some extent compared their scores with those of others. This finding suggests that our senior participants might not be familiar with the game elements for the first time when playing the games for they might be too occupied by the exercise tasks.

Finally, most reported that what motivates them to use the system are the participation in the exercise and the shared topics arising from the gameplay. It was observed that the game stimulates a great deal of social interaction. Throughout the whole session, all the participants were attentive to the gaming and how one or others could manage to perform better on the tasks. Most of them were practicing before they played the game.

9.5.4.3 Intent to Use

Most participants showed continuous intent to use EFS. Despite a few participants reporting that they found using the system troublesome, most found it beneficial in that they knew EFS was designed to improve their balance and they can exercise systematically by simply following it. Other benefits include that EFS would provide interactive features good for exercise, help maintain a regular exercise, and serve a social purpose of getting the family to exercise together.

9.6 Discussion

Our study first addressed the potential use of motion-sensing technology as a balance prevention tool for older people, identified the characteristics of designing a suitable gaming system, and then reported our findings from a participatory design session. The questionnaire interview on the suitability of the six embedded exercise shows that senior participants have varying ability of performing the six movements. While some might find it easy, others might find it hard. To deal with this issue, an adaptive mechanism can be designed to probe the range of motion in an individual senior at the beginning. Data obtained can then be used to build personal profiles, based on which the system will start with for subsequent gameplay. As the performance improves, the range of motion can gradually become more challenging.

The questionnaire interview on senior users' experience of using EFS shows that senior participants hold positive attitudes toward the use of EFS. Although they appeared to need some time to adapt to the motion-sensing interfaces, they were intrigued by doing exercise through gameplay. In line with previous findings, seniors value the game content that is meaningful and educational (Schutter and Vandenabeele 2008). When it comes to agility progress report, they appeared to pay little attention on it at first and then became aware of what it meant to them later on. This might be that they were so focused on the exercise itself rather than the information provided.

We believe that as they become more experienced, they would pay more attention on the game results. We might also redesign the progress report to make it more informative, such as provision of diagnostic report using the age group normative fitness scores and exercise advice from physiotherapists. It was also observed that gameplay can greatly facilitate social interaction. As was found in previous studies, senior participants enjoyed playing exergames with others rather than alone (Aarhus et al. 2011). Thus, the gameplay can be designed to enhance the effects of social interaction by including shared functions or interactive components between players.

Finally, the Activity Enjoyment Scale shows that older participants have positive emotions about EFS. As exercise enjoyment can be a predictor of participation in physical activity (Mullen et al. 2011), it is expected that participants will maintain exercise using EFS when this system is implemented in the intervention experiment.

We will also modify the game design by integrating two of the four game dynamics proposed by Priebatsch (2010), which are "influence and status dynamic" and "communal discovery dynamic." These are the dynamics that game designers leverage to make games fun and attractive. First, in game design, influence and status dynamic allows player to display one's ability to modify others' actions through social influence, such as the holding of digital trophy that represents status in achieving a certain game level. In the case of EFS, this dynamic can be used to encourage not only self-progress but also others' performance. For example, EFS can automatically videotape a few exemplary players based on the precision of the movements completed and then play their movements for others to emulate their performance.

Second, communal discovery means everyone work together to achieve a mutual goal. For example, collaborative games can be designed to involve individual's performance and use everyone's scores as a whole to level up. The agility scores of the entire group can be used to level up so all senior participants should practice hard to improve their agility. Note that we do not attempt to create a competing atmosphere as this might have a demotivating effect (Aarhus et al. 2011).

Since the aim of this study was to improve the design of EFS, the findings might not be comprehensive. Two issues arising from the study might provide directions for future studies. First, how the system can maximize the perceived benefits by the senior users should be explored. Other than the progress report coming from game results, there might be other components that can be taken as perceived benefits, such as diagnostic report from the system, physiotherapists' feedback, or senior fitness report. Second, it appears that this home-based exercise system works well for seniors when all participants are present in the same setting. Would the user experience remain positive when each individual is using the system at home? Future studies should examine how to maximize the perceived benefits and how individual or group mode would affect user experience.

9.7 Conclusion

EFS was developed to train senior adults' balance. Different from commercial products, our study adopted six exercises, which were specifically designed by the field experts to increase lower body strength of senior adults, and integrated them into video games. Using a participatory design, this study found that the senior participants enjoyed the game-based exercises and showed a positive experience using EFS. They provided insightful feedback on improving the system design regarding the suitability of the six exercises, system operation, game design, and intent to use.

It was found that (1) the system should include navigation requiring less learning, corrective feedback, timely prompts while idle; (2) game should be more diverse and relevant to their life, provide different levels or complexity of exercise tasks and make the scoring or progression report more explicit; and (3) what motivates them to use the system continuously appears to be interactive features, physical involvement, and social reasons. Moreover, the system can include an adaptive mechanism, which is capable of adjusting the motion recognition according to individual differences.

As part of an ongoing project, future study will implement EFS in an intervention program to evaluate the outcomes of the balance performance. We hope to provide more empirical evidence in the use of motion-sensing technology to improve the physical health of senior adults, as little research has thus far examined its effectiveness.

Acknowledgments This research was supported by the National Science Council, Taiwan, under project numbers NSC 102-2911-I-110-501, NSC 101-2511-S-110-003-MY3, NSC 100-2511-S-110-001-MY3, and NSC 99-2511-S-110-004-MY3.

References

Aarhus, R., Grönvall, E., Larsen, S. B., & Wollsen, S. (2011). Turning training into play: Embodied gaming, seniors, physical training and motivation. *Gerontechnology, 10*, 110–120. doi:10.4017/gt.2011.10.2.005.00.

Billis, A., Konstantinidis, E. I., Ladas, A. I., Tsolaki, M. N., Pappas, C., & Bamidis, P. D. (2011). Evaluating affective usability experiences of an exergaming platform for seniors. In: *Proceedings of the 10th International Workshop on Biomedical Engineering* (pp. 1–4). doi:10.1109/IWBE.2011.6079048.

Chao, K.-J., Huang, H.-W., Fang, W.-C., & Chen, N.-C. (2013). Embodied play to learn: Exploring Kinect-facilitated memory performance. *British Journal of Educational Technology*. doi:10.1111/bjet.12018.

Dishman, R. K. (1994). Motivating older adults to exercise. *Southern Medical Journal, 87*(5), S79–S82.

Doyle, J., Bailey, C., Dromey, B., & Scanaill, C. N. (2010, March). An interactive technology solution to deliver balance and strength exercises to older adults. *Paper Presented at the 4th International Conference on Pervasive Computing Technologies for Healthcare*. doi:10.4108/ICST.PERVASIVEHEALTH2010.8881.

Gerling, K., Livingston, I., Nacke, L., & Mandryk, R. (2012). Full-body motion-based game interaction for older adults. *Paper Presented at the 2012 ACM Annual Conference on Human Factors in Computing Systems*, Austin, Texas, USA. doi:10.1145/2207676.2208324.

Ijsselsteijn, W., Nap, H. H., Kort, Y. de, & Poels, K. (2007). Digital game design for elderly users. *Paper Presented at the 2007 Conference on Future Play*, Toronto, Canada. doi:10.1145/1328202.1328206.

Kahlbaugh, P. E., Sperandio, A. J., Carlson, A. L., & Hauselt, J. (2011). Effects of playing Wii on well-being in the elderly: Physical activity, loneliness, and mood. *Activities, Adaptation and Aging, 35*(4), 331–344. doi:10.1080/01924788.2011.625218.

Lange, B., Flynn, S., & Rizzo, A. (2009). Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation. *Physical Therapy Reviews, 14*(5), 355–363. doi:10.1179/108331909X12488667117258.

Maillot, P., Perrot, A., & Hartley, A. (2012). Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. *Psychology and Aging, 27*(3), 589–600. doi:10.1037/a0026268.

Marin, J. A., Felix, N. K., & Lawrence, E. (2011). Serious games to improve the physical health of the elderly: A categorization scheme. *Paper Presented at the Fourth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services*.

Mullen, S. P., Olson, E. A., Phillips, S. M., Szabo, A. N., Wójcicki, T. R., Mailey, E. L., et al. (2011). Measuring enjoyment of physical activity in older adults: Invariance of the physical activity enjoyment scale (paces) across groups and time. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1), 1–9.

Nelson, M. E., Rejeski, W. J., Blair, S. N., Duncan, P. W., Judge, J. O., King, A. C., et al. (2007). Physical activity and public health in older adults: Recommendation from the American College of Sports Medicine and the American Heart Association. *Medicine Science Sports Exercise, 39*(8), 1435–1445. doi:10.1249/mss.0b013e3180616aa2.

Priebatsch, S. (2010, August). Building a game layer on top of the world [Video file]. Retrieved from TED website: http://www.ted.com/talks/seth_priebatsch_the_game_layer_on_top_of_the_world.html

Raedeke, T. D. (2007). The relationship between enjoyment and affective responses to exercise. *Journal of Applied Sport Psychology, 19*(1), 105–115. doi:10.1080/10413200601113638.

Schutter, B. D., & Vandenabeele, V. (2008, May). Meaningful play in elderly life. *Paper presented at the 58th Annual Conference of the International Communication Association* "*Communicating for Social Impact*," Quebec, Canada.

Schutzer, K. A., & Graves, B. S. (2004). Barriers and motivations to exercise in older adults. *Preventive Medicine, 39*(5), 1056–1061. doi:10.1016/j.ypmed.2004.04.003.

Sherrington, C., Lord, S. R., & Finch, C. F. (2004). Physical activity interventions to prevent falls among older people: Update of the evidence. *Journal of Science and Medicine in Sport, 7*(1), 43–51. doi:10.1016/S1440-2440(04)80277-9.

Sheu, F.-R., & Chen, N.-S. (2014). Taking a signal: A review of gesture-based computing research in education. *Computers & Education, 78*, 268–277. doi:10.1016/j.compedu.2014.06.008.

Williams, M., Soiza, R., Jenkinson, A., & Stewart, A. (2010). EXercising with computers in later life (EXCELL)—pilot and feasibility study of the acceptability of the Nintendo(R) WiiFit in community-dwelling fallers. *BMC Research Notes, 3*(1), 238. doi:10.1186/1756-0500-3-238.

Chapter 10
Story-Based Virtual Experiment Environment

Ming-Xiang Fan, Rita Kuo, Maiga Chang and Jia-Sheng Heh

Abstract Science education is important for students developing critical thinking skills. How to engage students in science learning becomes an important issue. This research aims to design a virtual laboratory for students practicing science process skills. Moreover, the virtual laboratory uses story to raise students' learning motivation and uses a problem-solving procedure to navigate students while solving scientific problems. To verify the usefulness of the virtual laboratory, this research has conducted an experiment with 31 grade-10 participants. The experiment results show that students believe the virtual laboratory is useful for them in learning Physics. Since students' past Physics grades do not influence their intention of using the virtual laboratory, the story-based virtual laboratory can help both of low and high academic students in science learning and assist them establish their science process skills.

Keywords Science education · Science process skills · Model for problem-solving activity · Virtual experiment · Virtual laboratory · Storytelling

10.1 Research Motivation and Contribution

Laboratory plays an important role of science education for engaging the processes of investigation and inquiry (Hofstein and Lunetta 1982, 2004). Many teachers focus on improving students' high-stakes assessment performance in

M.-X. Fan · J.-S. Heh
Department of Information and Computer Engineering,
Chung-Yuan Christian University, Chung-Li, Taiwan

R. Kuo (✉)
Academic and Industrial Research Centre,
Knowledge Square, Inc., Kaohsiung, Taiwan
e-mail: rita.mcsl@gmail.com

M. Chang
School of Computing and Information Systems,
Athabasca University, Edmonton, Canada

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_10

terms of listening to lecture and memorizing the facts (Banilower et al. 2008). Students might have lower interests in learning science over time because of the boring lessons. Digital games could be a potential solution of increasing students' motivation in science learning (Honey and Hilton 2011).

Digital games are popular in the modern society. In USA, there are more than 180 million people playing digital games for more than 13 h per week (McGonigal 2011). To adapting the teaching methods for the students grown up with games, Prensky (2001) has suggested that digital games could be used as educational tools. Evidences show that games could enhance students' learning motivation (Cheng et al. 2012; Yang et al. 2012). Narrative context, goals, and rewards are important factors to attract students playing educational games (Dondlinger 2007).

This research aims to design a virtual laboratory, called virtual experiment environment, which combine the gaming factors and the virtual experiment to engage students in science learning. To reach the goal, the skills students need to learn from the virtual laboratory are considered. Moreover, to make students learn particular skills in the virtual laboratory, automatic navigation steps generation are designed.

This research first introduces important educational theories in science education in Sect. 10.2. Section 10.3 designs the story-based virtual experiment environment. The system is introduced in Sect. 10.4. Section 10.5 describes the research hypotheses and the experiment design. The collected data are analyzed and the findings are discussed in Sect. 10.6. Section 10.7 concludes the research and talks the possible future works.

10.2 Literature Review

10.2.1 Knowledge Structure

Knowledge structure is used for presenting the relationships among knowledge objects (Merrill 2000). It can be used for knowledge management (Cayzer 2004; Liu and Lee 2013), misconception diagnosis (Chen, 2011; Chang et al. 2003), and item generation (Gierl et al. 2012; Kuo and Chang 2009) in the area of advanced learning technology. Concept map, mind map, and conceptual diagrams are knowledge structures that are widely used for knowledge construction and sharing (Eppler 2006).

Wu et al. (2008) have designed a context-awareness knowledge structure that has successfully been used to generate mobile learning activities (Lu et al. 2011). The context-awareness knowledge structure is a three-layer structure, including the domain, concept, and object layers.

- Domain Layer: stores and presents the relationships among knowledge domains, such as Physics or Chemistry. It can also present the locations for mobile learning in the authenticate environments such as museums.
- Concept Layer: organizes the concepts related to the domain knowledge in hierarchical form. Take Fig. 10.1, for example, the concept "kinetic energy transfer to potential energy" is a type of "kinetic energy conservation," so the "kinetic energy transfer to potential energy" is the child node of the concept "kinetic energy conservation."

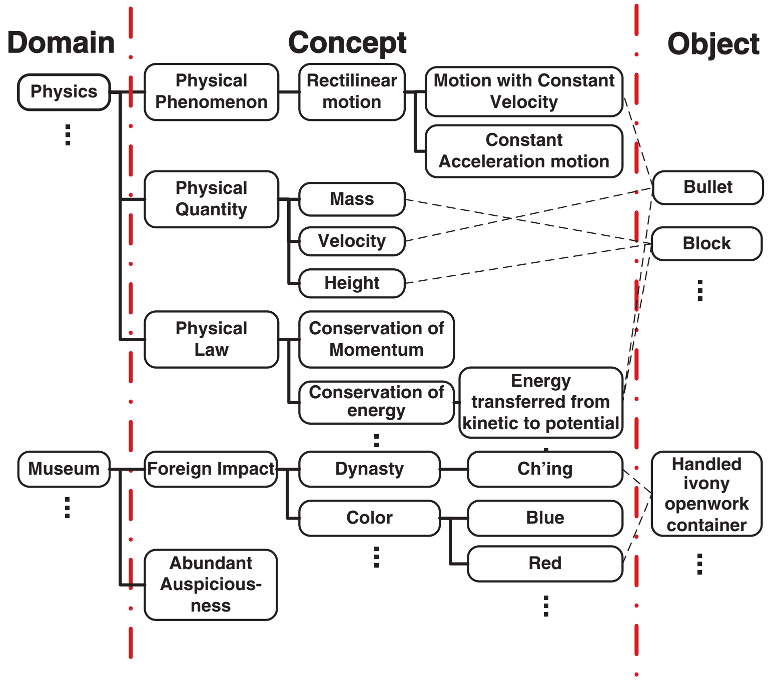

Fig. 10.1 Example of context-awareness knowledge structure

- Object Layer: stores the objects that particular domain knowledge in the concept layer may apply to. For example, the "bullet" is the object that can be used for explaining both of "motion with constant velocity" and "kinetic energy transfer to potential energy" concepts, is also related to the "velocity" and "height" Physics quantities while explaining those concepts.

Knowledge structure is used for storing the science concepts in this research. Most of these concepts are declarative knowledge. This research needs to further investigate what procedural knowledge students may need to learn in the virtual laboratories.

10.2.2 Science Process Skills and Virtual Laboratories

Scientists use science process skills to investigate the questions they met (Funk et al. 1979). Science process skills can help students form logical thinking (Koray and Koksal 2009) and develop cognitive functions (Ozgelen 2012). National Research Council (2009) also suggests that having science process skills should be one of science education goals.

Science process skills can be separated into basic and integrated science process skills. Basic science process skills are essential skill set in science

Table 10.1 List of science process skills (Funk et al. 1979)

Basic science skills	Integrated science process skills
Observation	Identifying variables
Classification	Constructing a table of data
Communication	Constructing a graph
Metric measurement	Describing relationships between variables
Prediction	Acquiring and processing data
Inference	Analyzing investigations
	Constructing hypotheses
	Defining variables operationally
	Designing investigations
	Experimenting

investigation process, such as observation and classification. Integrated science process skills are skills used in high-level problem-solving process—for instances, making hypotheses and analyzing collected data. Table 10.1 lists the science process skills proposed by Funk and colleagues in 1979.

Besides using traditional laboratory to train students' training science process skills, educators start to use technology to improve laboratory activities in science education. There are four major technologies applied in laboratory activities, including microcomputer-based laboratories, simulation/virtual reality, remote laboratories, and augmented reality (Chen et al. 2012). Considering the limitations of space, time, and budget, this research applies interactive simulation to develop the virtual laboratory environment.

Interactive simulation has been widely used in science education. Fan and Geelan (2012) have pointed out that interactive simulation has four major benefits for science education, which are sparking students' motivation (Tambade and Wagh 2011), supporting students' conceptual development (Fan and Geelan 2013), engaging students' inquiring skills (Mustafa and Trudel 2013), and realizing scientific discourse (Nichols et al. 2013). However, most of teachers are unfamiliar with programming and have difficulty to build their own interactive simulations for their courses (Sved 2010).

Easy Java Simulation (EJS) is one of the authoring tools for non-programmers, such as teachers, to create interactive simulations (EjsWiki n.d.). Take Physics, for example, researchers use EJS to teach harmonic motion, collision, geostationary orbit, etc. (Wee 2010). Hwang and Esquembre (2003) also have built more than 80 simulations for Physics phenomena. This research uses the simulations made by Hwang and Esquembre (2003) in the proposed virtual experiment environment.

10.2.3 Virtual Experiment Environment

Though the use of interactive simulations in virtual laboratories reduces the cost, researchers have found that it also has limitation on supporting the planning and designing activities such as asking students to identify or manipulate variables of a Physics phenomenon (Chen et al. 2012).

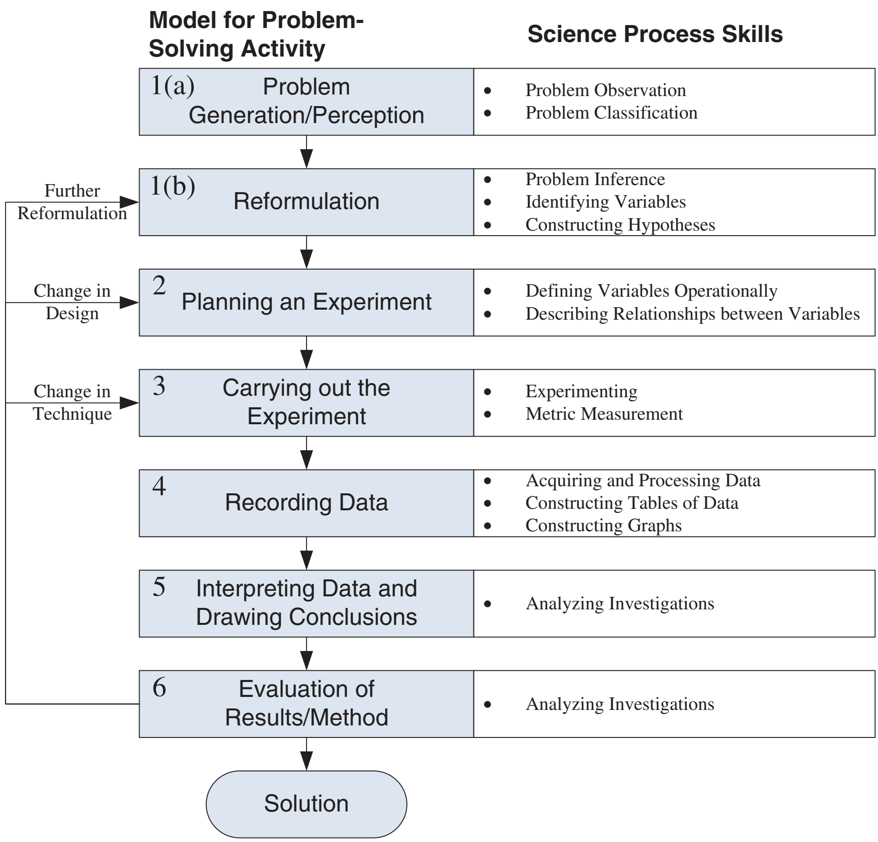

Fig. 10.2 Students' experiment flow in virtual experiment environment

Kuo et al. (2000) have designed a virtual experiment environment that takes the model for problem-solving activity (Gott and Murphy 1987; Welford 1986) into its design in order to guide students solving science problems. Each stage in the model of problem-solving activity helps students practice different science process skills. Figure 10.2 shows the process that students doing experiment in the virtual experiment environment and the science process skills that they are practicing at different stages.

At Stage 1(a), problem generation and perception, students see an animation in the beginning. They need to identify what kind of scientific problems the animation represents. Students can practice their observation and classification skills. Next, students need to find the independent and dependent variables in the identified problem at Stage 1(b). Students can practice their skills in terms of identifying variables and constructing hypotheses.

After the variables are identified, students can choose which apparatus, such as scale or timer, they need to use for doing the experiment at Stage 2. They can practice their skills in terms of deciding what they need to observe (such as time or weight), considering how to measure it (such as the use of timer or scale), and constructing the relationships among variables. At Stage 3, students have to choose which apparatus is suitable for a certain situation in the experimental process. For example, students have to pickup the scale with proper minimum unit (such as 1/4 inches or 1/8 inches) for measuring the distance travelled by the ball. They can practice their skills of experimenting and metric measurement at this stage.

At Stage 4, students need to collect the data in various forms, such as table, graph, etc. Acquiring and processing data, constructing tables of data, and constructing graphs are the skills they could learn at this stage. With the collected data, students can interpret the experiment results and make conclusion accordingly, at Stage 5. They can practice their skills of analyzing investigations at this stage.

When students cannot find the conclusion at Stage 5, the collected data might be inadequate for solving the problem. Students need to assess the methods and refine the experiment process at Stage 6. They can redefine the independent and dependent variables, choose other apparatuses, and use different techniques for data collection. Analyzing investigations is the skill they can practice at this stage.

10.3 Proposed Method

10.3.1 Story-Based Virtual Experiment Environment

This research integrates story element into the virtual experiment environment proposed by Kuo et al. (2000). When students use the system, the system first prompts a story and asks students to solve the problem involved in the story. For example, students role play an adventurer in a scene of the story, where they encounter a valley and have to find a way to cross the valley by using a vine to swing to the other side which is higher than the cliff at this side (as Fig. 10.3 shows). Students are at Stage 1(a) of the model for problem-solving activity at this moment.

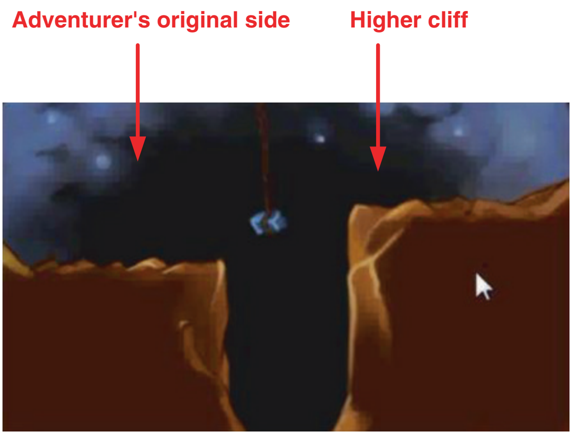

Fig. 10.3 Scenario of adventurer encountering a valley

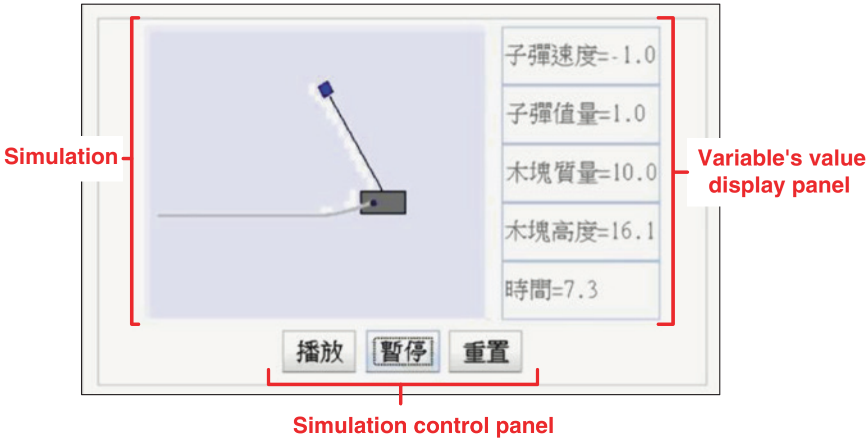

Fig. 10.4 Snapshot of the Java simulation (Hwang and Esquembre 2003)

Students then may make a hypothesis to solve this problem. For example, students can assume that when the adventurer runs faster, he or she may swing higher and reach the higher cliff at the other side of the valley. To verify the correctness of the hypothesis, students can choose one of the Java simulations (as Fig. 10.4 shows) developed by Hwang and Esquembre (2003) to preview the correspondent follow-up story scene when their assumption is applied.

Next, they have to find out which object in the story scene (such as the adventurer) is the corresponding object in the Java simulations (such as the bullet in a shooting simulation) as Fig. 10.5 shows.

They also need to consider which variables (such as mass or speed) they may need to operate and to observe in the simulation. Figure 10.6 lists the physical quantities; the chosen objects in the simulation have, such as speed, mass, friction, etc. Students can pickup the physical quantities from the right-hand side and move them to the left-hand side so the particular values can be observed and manipulated later. Students are at Stage 1(b) of the model for problem-solving activity while doing this.

At Stage 2, students need to consider what independent variable (which they need to manipulate), dependent variables (which they want to observe), and control variables (which are constant) are. In the above-mentioned story scene, students may think the running speed of the adventurer is an independent variable and the height that the rock can reach when it swings to the other side is the dependent variable. In Fig. 10.7, the variables selected from previous step, in Fig. 10.6, are listed on the right-hand side of the screen. Students can choose any of them and assign them to one of the independent, dependent, and control variable(s) blocks to complete the experiment plan.

Next, students run the simulation by adjusting the values of the independent variables and record the values of the dependent variables via observation. As

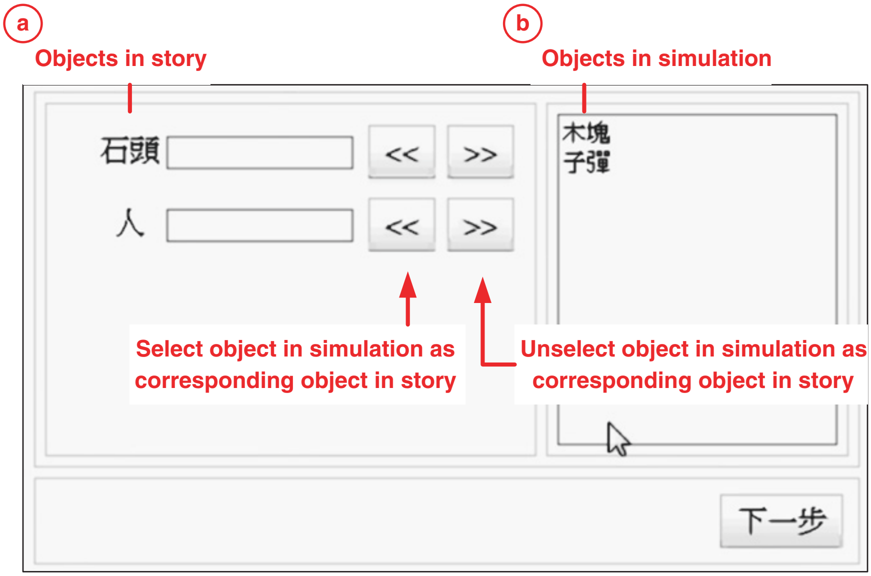

Fig. 10.5 Students need to match the objects in the simulator to the objects in the story animation

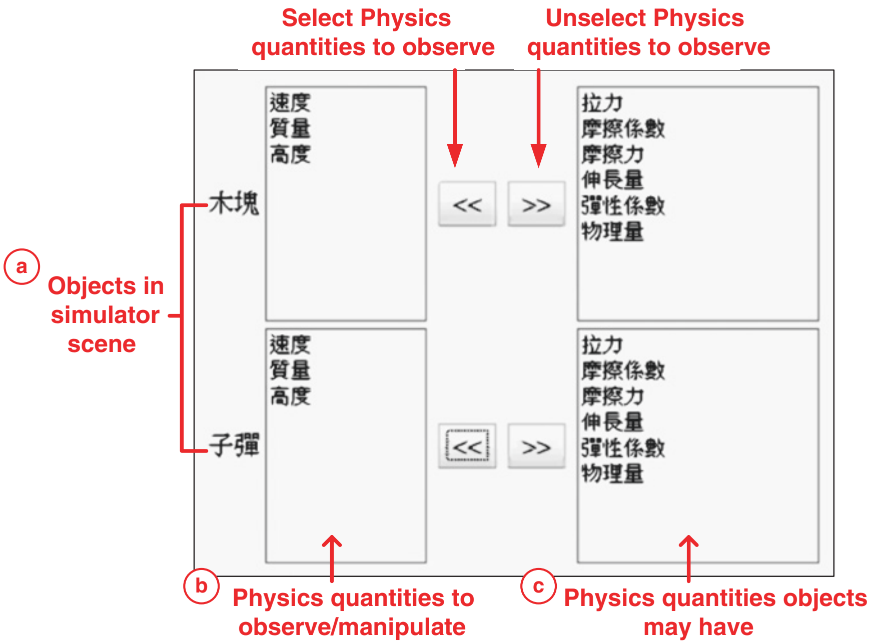

Fig. 10.6 Students need to consider which Physics quantities they would like to observe or control in the experiment

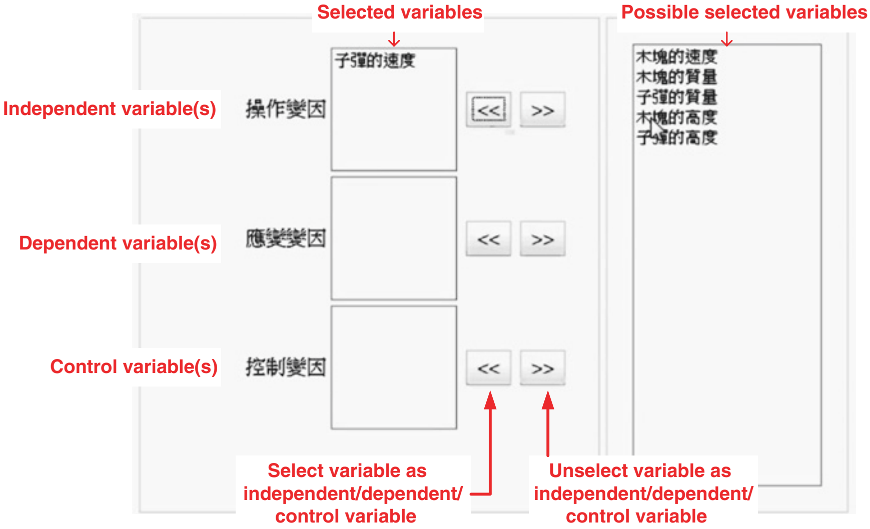

Fig. 10.7 Students need to select variables listed in the "possible selected variables" block and click " <<" buttons to assign the chosen variables to independent, dependent, or control block

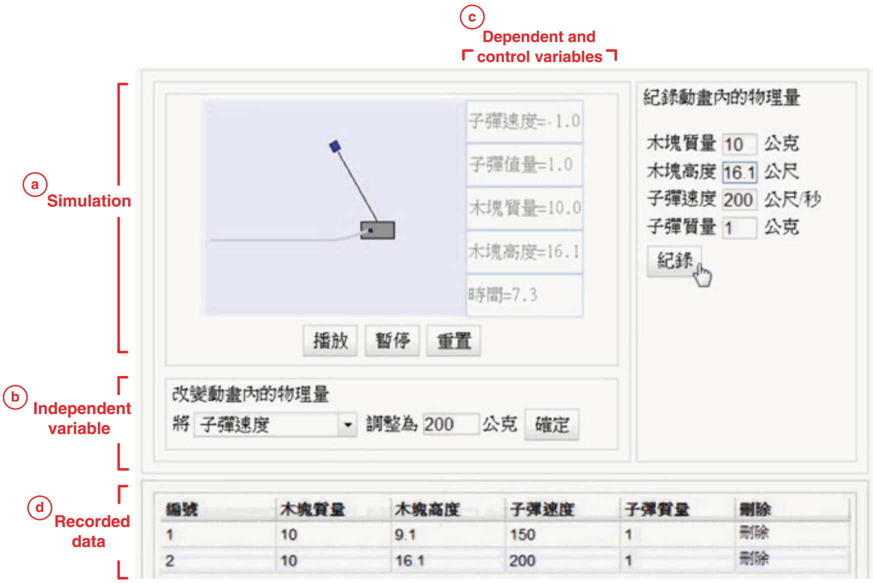

Fig. 10.8 Students run Java simulation (Hwang and Esquembre 2003) and record values of dependent variables via observation

Fig. 10.8 shows, students control Java simulation (in area ⓐ) via changing independent variable's value (in area ⓑ) and read the dependent variables' values (in area ⓒ). The recorded data is displayed in area ⓓ. When they do so in the virtual

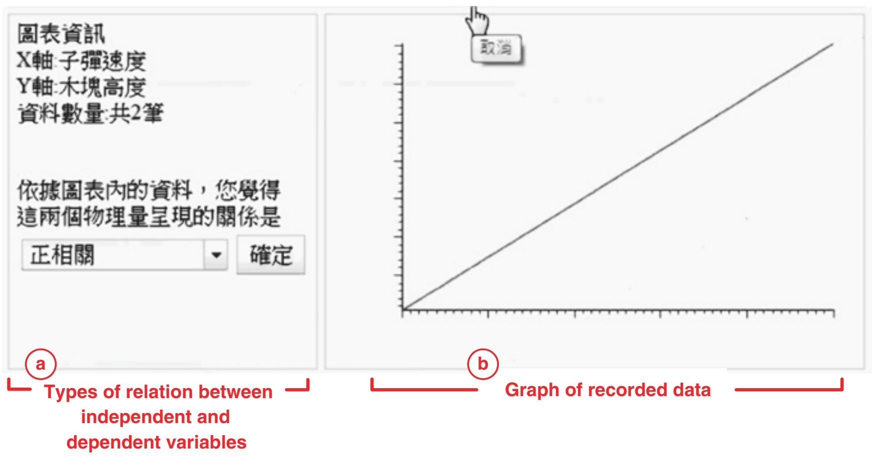

Fig. 10.9 Students observe the graph of recorded data and figure out the relation between the independent and dependent variable

experiment environment, they are at Stages 3 and 4 of the model for problem-solving activity.

After the experiment data are collected, students can examine the data and interpret the relations between the independent and dependent variables as they are at Stage 5. In Fig. 10.9, students can draw a graph according to the data recorded from the previous stage. Students also need to figure out which type of relation between independent and dependent variables to have, such as positive and negative correlation. At the end of the experiment, students need to assess the experiment results and to think whether or not the results confirm the hypothesis they made earlier. If the hypothesis is not supported by the results, students can go back to previous stages to refine their experiment design.

10.3.2 *Story-Based Context-Aware Knowledge Structure*

To automatically generate story and experiment scenes for each stage of the stages of model for problem-solving activity, this research designs a story-based context-aware knowledge structure to store the concepts in the scenes. The knowledge structure is modified from the context-awareness knowledge structure proposed by Wu et al. 2008). There are four layers in the knowledge structure as Fig. 10.10 shows.

The first layer is domain layer—stores the domain that the virtual experiment environment has experiments for students doing, such as Physics and Chemistry. The second layer is concept layer—stores the domain-relevant concepts in a

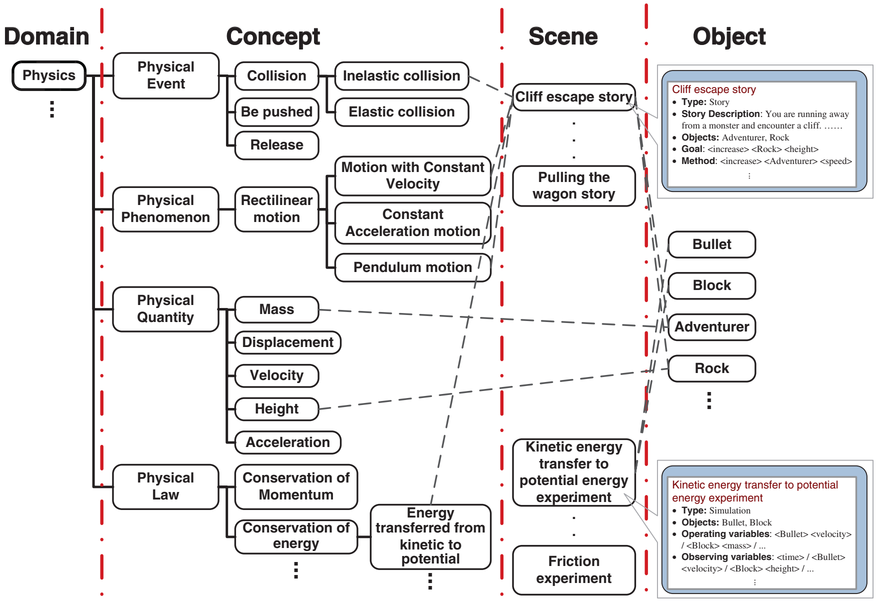

Fig. 10.10 Story-based context-aware knowledge structure example in physics

hierarchical structure. Taking Physics domain as example, the concept layer has four major concepts:

- Physical Event: is the main concept in the story and experiment scene, such as object collision.
- Physical Phenomenon: stores the physical phenomenon students can observe. For example, in rectilinear motion, an object may move in constant velocity (i.e., the phenomena of motion with constant velocity); when the object hits on another object hanging on the string, the pendulum movement may happen.
- Physical Quantity: is the physical attribute that an object has, such as its mass, velocity, acceleration, etc.
- Physical Law: is the principle of how physical quantities change in a class of phenomenon, such as Newton's second law and the law of conservation of energy (e.g., energy transferred from kinetic to potential).

The original third layer, object layer, proposed by Wu et al. (2008) is divided into two layers—scene layer and object layer—in this research. The scene layer stores the information of the story and its scene. Take "cliff escape" story scene as example, the node has a story description label, which is used for explaining to the students the scene in text; an object label that is used for associating the required objects in the object layer; the goal label that is used for the system to know what dependent variables and how their values should change to for solving the problem; and the method label stores the correspondent independent variables and how their values should change to for making the dependent variables' values solve the problem.

10.4 The System

A Physics example of energy transferred from kinetic to potential is demonstrated in this section. In the beginning, students can choose “cliff escape” story to start their learning as Fig. 10.11 shows.

The upper left of the screen shows the narration of the story animation that students can see at the upper right of the screen: *You are standing near a cliff and trying to reach the opposite side of the valley, which is higher than this side. Look! A rock is tied by a vine in the middle of the valley. Maybe you can jump on the rock and swing over to the other side. However, the other cliff is a little bit higher. How to swing over to the other side of the valley successfully?*

In the bottom of Fig. 10.11, students need to make a hypothesis about how to solve the problem. In this case, as the goal label in the cliff escape scene node stores, the goal of the problem in this story is to *increase the rock’s height* when it is swung to the other side of the valley; the method to reach the goal is to *increase adventure’s speed*. Students can choose the objects and its physical quantities from the drop-down list to make their hypothesis.

After making the hypothesis, students can select which simulation they would like to use to verify their hypothesis. They could watch the simulations and consider which simulation has similar situation to the story animation. In Fig. 10.12, students select one of the simulations from the simulation list (in area ⓐ) and click “Preview Simulation” button (in area ⓑ) that pops up a new window as area ⓒ shows. If students believe the chosen simulation fit the situation in the story, they can click “Next Step” button to go to the next stage.

Fig. 10.11 Students read the story description (in area ⓐ) and determine how to solve the problem. In this case, students think the goal of the problem (in area ⓒ) is *increasing rock’s height* and decide to *increase adventurer’s speed* to reach the goal (in area ⓑ)

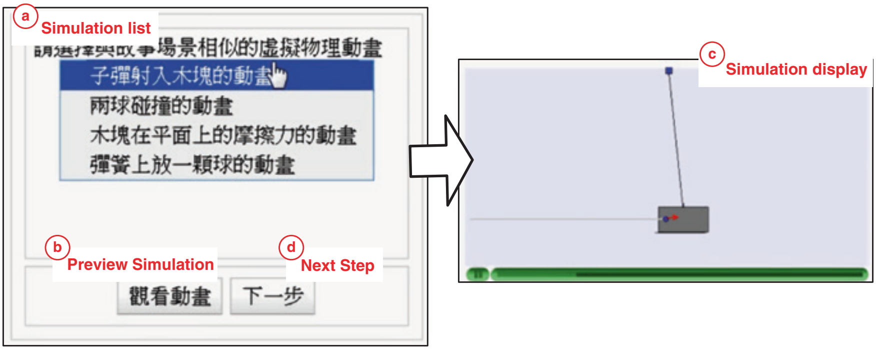

Fig. 10.12 Students can preview a Java simulation (Hwang and Esquembre 2003) before deciding which simulation they would like to use to verify their hypothesis

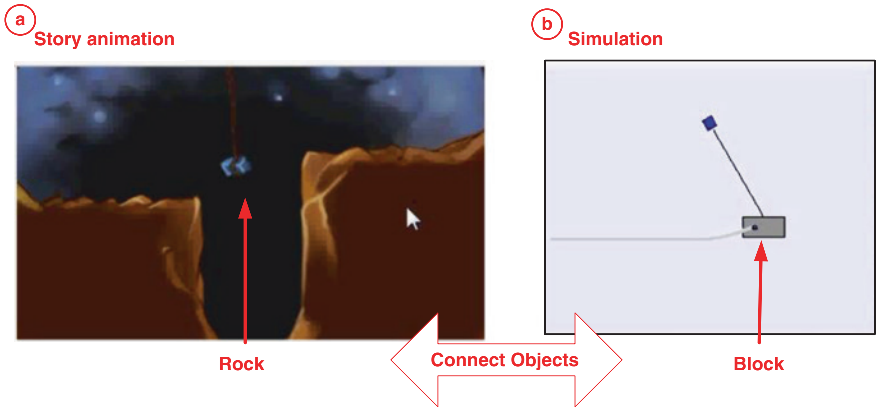

Fig. 10.13 Students can match objects in story animation to the simulation by previewing the chosen Java simulation (Hwang and Esquembre 2003)

Next, students have to connect the objects appear in both of the story animation and the chosen simulation together so they can apply the science concepts they have learnt and start their experiment. In the "cliff escape" example as area ⓐ in Fig. 10.13 shows, students have to find that the rock in the story animation in fact can be seen as the block in "A bullet shooting in a block tying under the pendulum" simulation (in area ⓑ in Fig. 10.13) and the adventurer in the story animation actually is the bullet in that simulation. Students also need to considers which physical quantities they may want to observe or manipulate later and decide which variable belongs to independent variable (which they need to manipulate) and which are dependent variables (which they want to observe).

Once students finish the settings for chosen variables, the system shows students again the simulation they chose earlier but this time some control panes are also provided for students doing experiment and watching the results. Area ⓐ in Fig. 10.14 is the control pane; students can play or pause the simulation at any

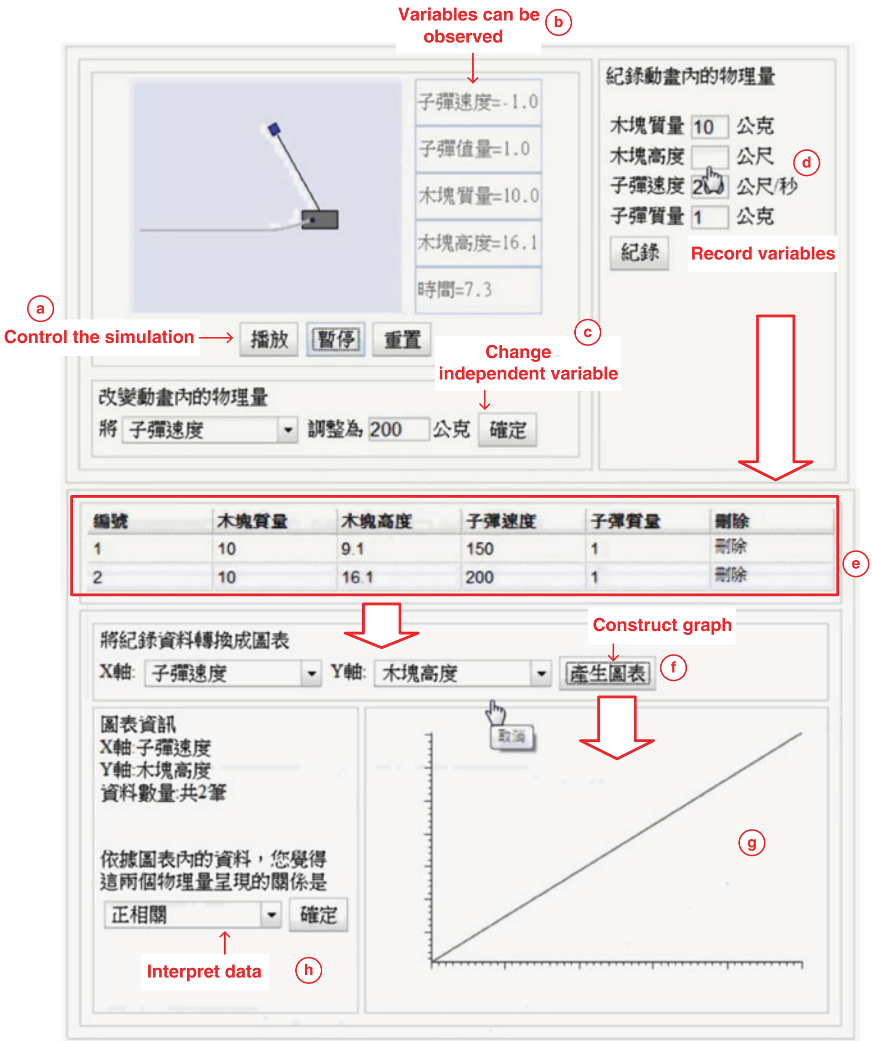

Fig. 10.14 Students can control the independent variable in the chosen simulation as well as observe and record the values of dependent variables. After all data are recorded, students can generate a graph to interpret the relation between independent and dependent variables

time they want and observe variables' values in area ⓑ. They could also change independent variable's value (e.g., change the bullet's speed from 100 m per second to 200) in area ⓒ and see the changes of the dependent variable's value through replays. Students need to manually record the data they observed from the simulation from time to time in area ⓓ, and the data will be added to the table in area ⓔ. After data are collected, students can generate a graph by clicking the button in area ⓕ and decide the variables for being x- and y-axes. The generated graph will be shown in area ⓖ.

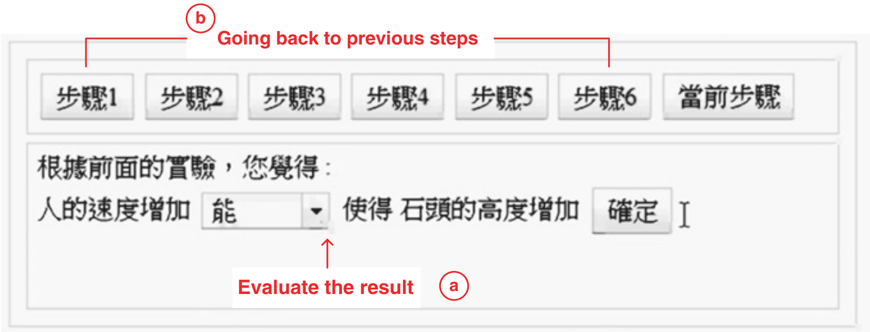

Fig. 10.15 Students can evaluate their hypothesis made at the first stage and see whether or not they need to re-plan or redo the experiment

Students can check the generated graph and consider the relations in-between the two variables; for instance, does the bullet's speed have positive (or negative) correlation with the height that the block will be hit to? In the case that Fig. 10.14 shows, students thought bullet's speed is independent variable and forms x-axis and thought the block's height is dependent variable and forms y-axis. Students can choose what kind of relations the two variables have in area ⓗ.

At the end, students need to assess the hypothesis they made earlier according to the experiment results as Fig. 10.15 shows. In this "cliff escape" example, students make a hypothesis—"when the adventurer runs faster and jumps to the rock, he or she may swing higher and reach the higher cliff at the other side of the valley." They further think the situation in the "A bullet shooting in a block tying under the pendulum" simulation is similar to the story animation and can be used to solve the problem they have encountered in the story. They believe that the bullet in the simulation can be seen as the adventurer in the story animation and the block can be seen as the rock. Since the experiment results prove that the bullet's speed and the block's height have positive correlation, increasing adventurer's speed can make the rock gets to higher place when it is swung to the other side of the valley. Their hypothesis is confirmed. If the experiment results cannot support the hypothesis they made, they can go back to any previous steps listed in the area ⓑ to re-plan and redo the experiment.

10.5 Research Design

10.5.1 Hypotheses and Experiment Design

We have several hypotheses for the proposed story-based virtual experiment environment:

- H1: Story-based virtual experiment environment is useful for students.
- H2: Gender will affect students' acceptance of using the system.

- H3: Students' past Physics grade will affect their acceptance of using the system.
- H4: Students' science process skills will affect their acceptance of using the system.
- H5: Students' past Physics grade is related to their science process skills.

To understand students' acceptance of using the system, this research adopts the technology acceptance model (TAM) from previous research (Davis et al. 1989) in which twelve 5-point Likert questions for three factors—perceived usefulness (PU), perceived ease of use (PEU), and behavior intention (BI) are designed.

Additionally, this research adopts test of integrated science process skills (TIPS) (Monica 2005) in which 30 items developed to assess students' science process skills. In the TIPS, there are 19 items related to the four skills, which students are going to practice in the proposed system:

- 6 items belong to the skills of stating hypotheses which is corresponding to Stage 1(b).
- 3 items belong to the skills of operational definitions which is corresponding to Stage 1(b).
- 7 items belong to the skills of identifying and controlling variables which is corresponding to Stage 2.
- 3 items belong to the skills of graphing and interpreting data which is corresponding to Stage 4.

Figure 10.16 shows the relations among factors that are involved in the hypotheses.

An experiment was conducted in a high school class in Tainan City, Taiwan, to verify the hypotheses. There were 31 grade-10 participants (23 males and 8 females) and they all learned basic Dynamics before. In the beginning, participants took 50 min to write their TIPS. After that, the researchers taught them how to use the story-based virtual experiment environment. The system had four story scenes and participants took next 50 min to use the system and to solve the problems by

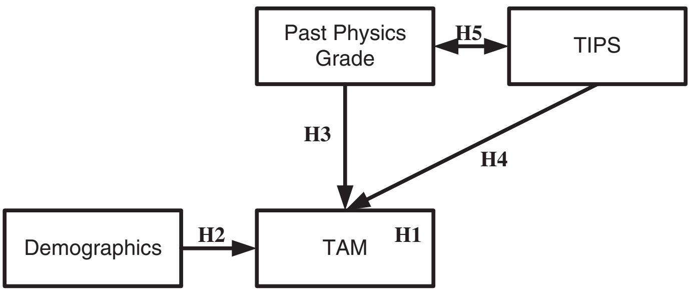

Fig. 10.16 Relations among factors involved in the hypotheses

conducting their own experiments. At the end, participants were asked to write a questionnaire.

10.5.2 Reliability and Validity

Before we verified the reliability and validity of the collected data, we found 13 responses were inconsistent in terms of answering the flip-flop items I_6 and I_8. After the removal of the inconsistent responses, we used SPSS 17.0 to test the reliability and validity of the data. Based on the test results, the TAM questionnaire was eliminated to 10 questions for the three factors. The reliability of the TAM was acceptable as its Cronbach's alpha value is 0.823 (George and Mallery 2010). Next, we used principle component analysis to test the validity of the TAM. As the results listed in Table 10.2 show, the data collected are valid.

Table 10.2 Validity analysis results of the technology acceptance model

Item	Factor		
	1	2	3
Factor 1: perceived usefulness			
I_2: Using story-based virtual experiment environment would improve my manipulating abilities in Physics experiment	0.928		
I_5: Using story-based virtual experiment environment would make me understand the problem in the story scenes	0.801		
I_1: I would find story-based virtual experiment environment is useful to me	0.766		
I_3: Using story-based virtual experiment environment would teach me how to find the relation between physics quantities	0.748		
Factor 2: behavior intension			
I_{19}: I would like to the story-based virtual experiment environment in the future.		0.799	
I_{20}: I would like to introduce story-based virtual experiment environment to others.		0.759	
I_{21}: I would like to use systems similar to the story-based virtual experiment environment.		0.678	
Factor 3: perceived ease of use			
I_6: I would find story-based virtual experiment environment's user interface easy to use.			0.902
I_7: Learning to operate story-based virtual experiment environment would be easy for me			0.693
I_8: I frequently could not find the function which I would like to use in story-based virtual experiment environment.			0.658
Eigenvalue	4.322	1.688	1.364
% of variance	43.334	16.883	13.642
Overall $\alpha = 0.823$, total variance explained is 73.748 %			

10.6 Evaluation and Discussion

10.6.1 Findings

Eighteen participants' data, including 14 males' and 4 females' data, were used for the evaluation after the removal of the inconsistent responses. Their average past Physics grade is 73.28 (male: 74.79; female: 68.00), and the average TIPS score is 70.11 (male: 69.68; female: 71.00). There are no significant difference between male and female participants in terms of their past Physics grade and TIPS score (past Physics grade: $t = 1.258$, $p = 0.226$; TIPS score: $t = -0.197$, $p = 0.847$). Table 10.3 shows the descriptive statistics of the participants.

The descriptive statistics of the TAM is listed in Table 10.4. The average score of their technology acceptance level is 3.53 (between Neutral and Agree). The average score for Perceived Usefulness is 3.74, which means, students believe the story-based virtual experiment environment can help them learning Physics concepts. Therefore, hypothesis H1 is supported. However, the average scores for PEU and BI are not high enough and students' perceptions toward the easy of use and the intention of using the proposed system are between Disagree and Neutral.

There is significant difference between male and female participants' acceptance level ($t = -2.169$, $p = 0.045 < 0.05$) as the data listed in Table 10.5 show, which shows female participants accept the use of the proposed system much more than male participants. Therefore, hypothesis H2 is also supported.

We also find that there is a significant positive correlation between PU and BI factors as the results listed in Table 10.6 show. This finding shows that students' intention of using an educational system will be increased if they think the system is useful.

We also want to know whether or not students' past Physics grades and TIPS scores will affect their acceptance of using the proposed system. The Pearson correlation test shows that there is no significant correlation between students' past Physics grade and TAM factors as well as between students' TIPS scores and

Table 10.3 Descriptive statistics in gender, past Physics grade, and TIPS score

Gender	No. of Participants	Past physics grade		TIPS score	
		Mean	SD	Mean	SD
Male	14	74.79	9.553	69.86	9.968
Female	4	68.00	9.345	71.00	11.431
All	18	73.28	9.676	70.11	9.964

Table 10.4 Descriptive statistics of participants' technology acceptance

Gender	No.	PU		PEU		BI		Overall	
		M	SD	M	SD	M	SD	M	SD
Male	14	3.59	0.67	2.76	1.00	2.76	1.00	3.38	0.57
Female	4	4.25	0.87	3.75	0.32	3.75	0.32	4.08	0.55
All	18	3.74	0.74	2.98	0.98	2.98	0.98	3.53	0.63

PU: perceived usefulness; PEU: perceived ease of use; BI: behavior intention

Table 10.5 Independent t-test result for different genders' acceptance levels toward the system

		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Sig (2-tailed)
PU	Equal variances assumed	1.430	0.249	−1.640	16	0.120
	Equal variances not assumed			−1.410	4.084	0.230
PEU	Equal variances assumed	0.726	0.407	−1.040	16	0.314
	Equal variances not assumed			−1.117	5.414	0.311
BI	Equal variances assumed	2.655	0.123	−1.912	16	0.074
	Equal variances not assumed			−3.176	15.422	0.006
Overall	Equal variances assumed	0.229	0.639	−2.169*	16	**0.045**
	Equal variances not assumed			−2.215	5.015	0.077

PU: perceived usefulness; PEU: perceived ease of use; BI: behavior intention
*: $p < 0.05$ (2-tailed)

Table 10.6 Correlation analysis among TAM factors

	PU versus PEU	PU versus BI	PEU versus BI
Pearson correlation	0.201	**0.557***	0.273
Sig. (2-tailed)	0.425	0.016	0.273
N	18	18	18

PU: perceived usefulness; PEU: perceived ease of use; BI: behavior intention
*: correlation is significant at the 0.05 level (2-tailed)

TAM factors. Furthermore, we conduct another round of correlation analysis for the TIPS 19 items whose associated skills can be practiced in the proposed system against to the three TAM factors. The results still show no significant correlation between the four skills and TAM factors. Table 10.7 lists all the analysis results of the Pearson correlation test. We also find that there is no significant relation between students' past Physics grades and TIPS scores ($r = 0.289$, $p = 0.244 > 0.5$). Therefore, the last three hypotheses are not supported.

10.6.2 Discussion

Although the average students' acceptance levels toward the proposed system are between Neutral and Agree, their PEU and intention of using the system are only between Neutral and Disagree. We interviewed some students and wanted to know why they do not think the system is ease of use. One of the reasons is they cannot find the functions they want to use at very beginning; hence, giving students enough time to get themselves familiar with the system before starting, the pilot-relevant activities should be considered for future pilots. Another reason is that the students

Table 10.7 Correlation analysis between students' tests scores (including past physics grade and TIPS score) and TAM factors

		PU	PEU	BI	TAM
Past physics grade	Pearson correlation	−0.124	−0.404	−0.012	−0.214
	Sig. (2-tailed)	0.624	0.097	0.963	0.395
	N	18	18	18	18
TIPS score	Pearson correlation	0.448	−0.076	−0.106	0.135
	Sig. (2-tailed)	0.062	0.763	0.675	0.592
	N	18	18	18	18
TIPS_Var Score	Pearson correlation	0.444	0.089	0.135	0.308
	Sig. (2-tailed)	0.065	0.729	0.592	0.213
	N	18	18	18	18
TIPS_Hyp Score	Pearson correlation	0.280	−0.173	−0.373	−0.106
	Sig. (2-tailed)	0.280	0.492	0.127	0.676
	N	18	18	18	18
TIPS_Def Score	Pearson correlation	0.359	−0.192	0.270	0.227
	Sig. (2-tailed)	0.144	0.445	0.278	0.364
	N	18	18	18	18
TIPS_Grp Score	Pearson correlation	−0.112	0.020	−0.408	−0.238
	Sig. (2-tailed)	0.657	0.937	0.093	0.341
	N	18	18	18	18

PU: perceived usefulness; PEU: perceived ease of use; BI: behavior intention; TAM: technology acceptance model; TIPS_Var Score: identifying and controlling variables score; TIPS_Hyp Score: stating hypotheses score; TIPS_Def Score: operational definitions score; TIPS_Grp Score: graphing and interpreting data score

feel the simulations (i.e., the EJS as Fig. 10.14 shows) too complicated to use than other steps, and students need more time to learn how to use the user interface of the EJS. Therefore, to make students have higher acceptance level toward the use of the system, simplifying the user interface of the simulation would be important.

On the other hand, female participants have higher acceptance in terms of using the system according to the t-test results listed in Table 10.5. This finding is similar to other studies on gender differences in learning technology research domain (Arbaugh 2000; Viberg and Gronlund 2013). Female participants may be more active in learning process than male participants (Gonzalez-Gomez et al. 2012) and may prefer storytelling games (Robertson 2012). Robertson's research result indicates that story-based educational game may engage female students' better and our research result is in line of it.

Female participants may also have higher intention of using the system because of the perceived social pressure, such as the pressure from teachers (Cheung and

Lee 2011). To verify this hypothesis, we further evaluate the responses of two social inference questions:

- I_{13}: I use the system because my teacher wants me to play it.
- I_{14}: I use the system because my teacher asks me to play it.

Although we find that there is no significant gender difference in the responses of the two questions (I_{13}: $t = -1.072$, $p = 0.300 > 0.05$; I_{14}: $t = -0.905$, $p = 0.379 > 0.05$), the correlation analysis results show that both questions significantly are in positive correlation to TAM (I_{13}: $r = 0.726$, $p = 0.001 < 0.05$; I_{14}: $r = 0.509$, $p = 0.031 < 0.05$). Teachers' attitudes toward the use of new technology for students learning significantly influence students' acceptance levels toward the new technology. Researchers should clearly explain to teachers the benefit that both teachers and students can receive through the use of their proposed educational systems (including educational games) and get teachers' support before applying their proposed systems and games into student learning process.

Our research result also shows that there is no significant correlation between students' grades (include past Physics grades and TIPS scores) and their technology acceptance levels. The finding indicates that the system can attract all students including both lower and higher academic achievement students. With teachers' encouragement, lower academic students may spend more time on using the system and practice their science process skills. When they find that their skills are improved, they may have more intention to actively use the system because the analysis result shows that PU has significantly positive correlation to Behavior Intention factor.

Although we find that there is no significant relation between students' past grades and TIPS scores, the reason may be—the traditional Physics lessons are not focusing on science process skills. Therefore, using story-based virtual experiment environment could fill in the missing puzzle piece that traditional science learning has and enhance students' science process skills.

10.7 Conclusion

This research designs a story-based virtual experiment environment for students practicing their science process skills. Students can improve their skills at every stages of model for problem-solving activities. These practices could help students solve the real-world problems through applying the knowledge they learned in the school.

The experiment result shows that students, especially for female students, believe that the system is useful for them in terms of learning. Since students' past Physics grades and TIPS scores would not influence their intention of using the system, we could design more attractive stories to engage both low and high academic students practicing their science process skills, which is usually not paid attention in traditional science learning.

After the removal of the inconsistent responses, this research only analyzes 18 students' intentions of using the proposed system. Collecting more data for the analysis should be considered in the future studies. Students' behaviors in the system should also be analyzed according to the science process skills at each stage.

References

Arbaugh, J. B. (2000). An exploratory study of the effects of gender on student learning and class participation in an Internet-based MBA course. *Management Learning, 31*(4), 503–519.

Banilower, E., Cohen, K., Pasley, J., & Weiss, I. (2008). *Effective science instructions: What does research tell us?*. Portsmouth: RMC Research Corporation, Center on Instruction.

Cayzer, S. (2004). Semantic blogging and decentralized knowledge management. *Communications of the ACM, 47*(12), 47–52.

Chang, J.-C., Chiu, Y.-P., Lin, Y.-Y., & Heh, J.-S. (2003). Using knowledge map to diagnose misconceptions in real-time internet tests. In D. Lassner & C. McNaught (Eds.). *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* 2003 (pp. 1297–1300). Chesapeake, VA: AACE.

Chen, L.-H. (2011). Enhancement of student learning performance using personalized diagnosis and remedial learning system. *Computers and Education, 56*(1), 289–299.

Chen, S., Lo, H.-C., Lin, J.-W., Liang, J.-C., Chang, H.-Y., Hwang, F.-K., et al. (2012). Development and implications of technology in reform-based physics laboratories. *Physical Review ST Physical Education Research, 8*(2), 020113.

Cheng, Y. M., Kuo, S. H., Lou, S. J., & Shih, R. C. (2012). The construction of and online competitive game-based learning system for junior high school students. *Turkish Online Journal of Educational Technology, 11*(2), 214–227.

Cheung, C. M. K., & Lee, M. K. O. (2011). Exploring the gender difference in student acceptance of an internet-based learning medium. In T. Teo (Ed.), *Technology acceptance in education* (pp. 183–199). Rotterdam: Sense Publishers.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

Dondlinger, M. J. (2007). Educational video game design: A review of the literature. *Journal of Applied Educational Technology, 4*(1), 21–31.

EjsWiki. (n.d.). Retrieved November 4, 2013, from http://www.um.es/fem/EjsWiki/pmwiki.php.

Eppler, M. J. (2006). A comparison between concept map, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information Visualization, 5*(3), 202–210.

Fan, X., & Geelan, D. (2012). Integration information technology and science education for the future: A theoretical review on the educational use of interactive simulations, *Proceedings of the 2012 Australian Computers in Education Conference*. Paper presented at Wesley College, Perth, Australia, Oct 2–5 2012. Retrieved November 4, 2013, from http://acec2012.acce.edu.au/sites/acec2012.acce.edu.au/files/proposal/133/MyRefereedPaper_0.pdf.

Fan, X., & Geelan, D. (2013). Enhancing students' scientific literacy in science education using interactive simulations: A critical literature review. *Journal of Computers in Mathematics and Science Teaching, 32*(2), 125–171.

Funk, H. J., Fiel, R. L., Okey, J. R., Jaus, H. H., & Sprague, C. S. (1979). *Learning science process skills*. Iowa: Kendall, Hunt.

George, D., & Mallery, P. (2010). *SPSS for windows step by step: A simple guide and reference 18.0 update* (11th ed.), Boston: Allyn & Bacon.

Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education, 46*(8), 757–765.
Gonzalez-Gomez, F., Guardiola, J., Rodriguez, O. M., & Alonso, M. A. M. (2012). Gender difference in e-learning satisfaction. *Computers and Education, 58*(1), 283–290.
Gott, R., and Murphy, P. (1987). *Assessment of performance unit. Science report for teachers: 9: Assessing investigations at ages 13 and 15*. London: Department of Education and Science.
Honey, M. A., & Hilton, M. (2011). *Learning science through computer games and simulations*. Washington, DC: National Research Council.
Hofstein, A., & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects of research. *Review of Educational Research, 52*(2), 201–217.
Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88*(1), 28–54.
Hwang, F.-K., & Esquembre, F. (2003). Easy java simulation: An interactive tool for conceptual learning of science. In D. Lassner & C. McNaught (Eds.). *Proceeding of World Conference on Educational Multimedia, Hypermedia and Telecommunications* 2003 (pp. 791–794). Chesapeake, VA: AACE.
Koray, O., and Koksal, M. S. (2009). The effect of creative and critical thinking based laboratory applications on creative and logical thinking abilities of prospective teachers, *Asia-Pacific Forum on Science Learning and Teaching, 10*(1), Article 2.
Kuo, R., & Chang, M. (2009). Elementary level botanical item generation. *IEEE Learning Technology Newsletter, 11*(3), 7–9.
Kuo, L.-P., Dong, D.-X., Hsu, C.-K., & Heh, J.-S. (2000). Design an enhanced virtual experiment environment using science process skills on WWW. In J. Bourdeau & R. Heller (Eds). *Proceedings of the AACE World Conference on Educational Multimedia, Hypermedia & Telecommunications* (p. 1820). Chesapeake, VA: AACE.
Liu, S.-H., & Lee, G.-G. (2013). Using a concept map knowledge management system to enhance the learning of biology. *Computers and Education, 68*(1), 105–116.
Lu, C., Chang, M., Kinshuk, Huang, E., & Chen, C.-W. (2011). Architecture and collaborations among agents in mobile educational game. In the *Proceedings of the 7th IEEE International Workshop on PervasivE Learning, Life, and Leisure, in IEEE Pervasive Computing and Communication 2011 (PerCom 2011)* (pp. 171–175), Seattle, WA, USA, 21–25 March 2011.
McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York: Penguin Press.
Merrill, M. D. (2000). Knowledge objects and mental models. *Proceeding of International Workshop on Advanced Learning Technologies 2000* (pp. 244–246). Los Alamitos: IEEE Computer Society Press.
Monica, K. M. M. (2005). *Development and validation of a test of integrated science process sills for the further education and training learners*. (Unpublished doctoral dissertation). University of Pretoria South Africa, Pretoria, South Africa.
Mustafa, M. I., & Trudel, L. (2013). The impact of cognitive tools on the development of the inquiry skills of high school students in physics. *International Journal of Advanced Computer Science Application, 4*(9), 124–129.
National Research Council. (2009). *Learning science in informal environments: people, places, and pursuits*. Washington, DC: The National Academics.
Nichols, K., Hana, J., & Ranasinghe, M. (2013). Transforming the social practices of learning with representations: A study of disciplinary discourse. *Research in Science Education, 43*(1), 179–208.
Ozgelen, S. (2012). Students' science process skills within a cognitive domain framework. *Eurasia Journal of Mathematics, Science & Technology Education, 8*(4), 283–292.
Prensky, M. (2001). Digital natives, digital immigrants: Part 1. *On the horizon, 9*(5), 1–6.
Robertson, J. (2012). Making games in the classroom: Benefits and gender concerns. *Computers and Education, 59*(1), 385–398.

Sved, J. A. (2010). Genetics computer teaching simulation programs: Promise and problems. *Genetics, 185*(4), 1537–1540.

Tambade, P. S., & Wagh, B. G. (2011). Assessing the effective ness of computer assisted instructions in physics at undergraduate level. *Eurasian Journal of Physics and Chemistry Education, 3*(2), 127–136.

Viberg, O., & Gronlund, A. (2013). Cross-cultural analysis of users' attitudes toward the use of mobile devices in second and foreign language learning in higher education: A case from Sweden and China. *Computers and Education, 69*(1), 169–180.

Wee, L. K. (2010). Physics educators as designers of simulation using easy java simulation (Ejs). *Proceeding of American Association of Physics Teachers Nation Meeting Conference: 2010 Summer Meeting*, Portland, 17–23 July 2010 (FI05).

Welford, G. (1986). *Science in schools: age 15, research report no. 4: a report on the 1983 survey of science performance of pupils of age 15 to the Department of Education and Science*. London: Department for Education and Science.

Wu, S., Chang, A., Chang, M., Liu, T.-C., & Heh, J.-S. (2008). Identifying Personalized Context-aware Knowledge Structure for Individual User in Ubiquitous Learning Environment, *Proceedings of the 5th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education* (pp. 95–99), Los Alamitos: IEEE Computer Society Press.

Yang, J. C., Chien, K. H., & Liu, T. C. (2012). A digital game-based learning system for energy education: An energy conservation pet. *Turkish Online Journal of Educational Technology, 11*(2), 29–37.

Chapter 11
Development of Handheld Augmented Reality X-Ray for K-12 Settings

Marc Ericson C. Santos, Mitsuaki Terawaki, Takafumi Taketomi, Goshiro Yamamoto and Hirokazu Kato

Abstract Mobile augmented reality is a next-generation interface for seamless ubiquitous learning. It offers many novel interactions that enable the visualization of digital information on real places and objects. These interactions require user-based testing for suitability in educational settings. One important interaction is augmented reality X-ray—providing an illusion to look inside objects. In this chapter, we implement augmented reality X-ray on a tablet computer by modifying the live video feed with computer graphics. Then, we evaluated our prototype based on the students' perception of depth, legibility, and realism. Results show that augmented reality X-ray hampers legibility. However, it does not have a significant impact on the perception of depth and realism. In our interviews with teachers, we found that augmented reality X-ray is perceived to be useful because it promotes learning by experience. It has the potential to improve both student attention and motivation. However, the teachers require a high-quality lesson plan, and extra training to use augmented reality X-ray in the classroom effectively.

Keywords Augmented reality X-ray · Contextual visualization · Mobile augmented reality · Prototype evaluation · User-based studies

11.1 Introduction

Many augmented reality systems refer to "X-ray vision" as the primary use of augmented reality. Sample applications include seeing through a pilot's cockpit floor and walls (Furness 1986), visualizing ultrasound information within a patient's body (Bajura et al. 1992), looking through buildings in local navigation (Sandor et al. 2010a) and studying underground pipes in construction (Zollmann

M.E.C. Santos (✉) · M. Terawaki · T. Taketomi · G. Yamamoto · H. Kato
Graduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan

M. Chang and Y. Li (eds.), *Smart Learning Environments*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44447-4_11

et al. 2010). Despite the many advances in prototypes, the realization of augmented reality X-ray remains challenging for general application because of the lack of user-based studies. Currently, researchers are studying various depth cues based on the current understanding of the human visual system, and ways of measuring perceived depth to evaluate usability (Livingston et al. 2013).

In educational settings, one important affordance of augmented reality is to "visualize the invisible" such as unobservable scientific concepts (Wu et al. 2012). For example, unseen forces acting on an object and magnetism have been illustrated using augmented reality. To teach the concept of force, Sotiriou et al. (2006) used augmented reality to integrate virtual arrows onto real carts found in a science museum display. Matsutomo et al. (2012) used augmented reality to draw virtual magnetic field lines onto real magnets. In their prototype, the shape of the magnetic field lines is computed in real time to demonstrate how two magnets affect each others. Aside from naturally invisible concepts, an object may be practically invisible due to occlusion such as internal organs and engines of machines. This can be addressed using augmented reality X-ray or providing the illusion of being able to peer inside a target object (Santos et al. 2013a). Augmented reality X-ray is achieved by modifying a live video feed of the object with computer graphics and occlusion cues.

Recently, researchers (Dey et al. 2012) have described methods for porting augmented reality X-ray to handheld devices making it practical for mobile augmented reality learning. Currently, applications such as Environmental Detectives (Klopfer and Squire 2008) and EcoMOBILE (Kamarainen et al. 2013) have recommended a new instructional method that uses mobile augmented reality games to investigate a real environment. According to these researchers, mobile augmented reality supports situated learning by facilitating interactions among learners and interaction with the real environment.

Augmented reality X-ray is another interaction with the environment that instructional designers can take advantage of in a location-based game, or other instructional designs. Using augmented reality X-ray, real objects found in an environment becomes a trigger for presenting information. Thus, augmented reality X-ray offers contextual visualization—the presentation of virtual information in the rich context of a real environment—thereby offering more compelling learning experiences.

11.1.1 Motivation

Augmented reality X-ray is a novel interaction for education. Thus, it is necessary to investigate how it affects the students' perception and the suitability of the state of the art to the teachers' practice. State-of-the-art implementations of handheld augmented reality X-ray have not yet been tested extensively.

Currently, the target is pedestrian applications such as navigation (Dey et al. 2012), thus the user-based studies focus on depth perception on medium-field (beyond arms-reach to 30 m) and far-field (beyond 30 m) distances. Implementations in these distances may not be suitable for the near-field (within arms-reach), which is the case when students approach a real object and try to observe its interior.

Moreover, current augmented reality X-ray methods use occlusion cues which sacrifices the legibility of the virtual object to convey depth. In educational settings, legibility is more important than conveying depth so that students can investigate the interior of a real object.

11.1.2 Contribution

We describe the implementation of an edge-based augmented reality X-ray using a tablet computer and evaluate it in the near-field distance (beyond arms-reach) with students and teachers for the first time. Our augmented reality X-ray interaction can be directly integrated into instructional designs based on mobile augmented reality such as location-based games, and other possible methods.

Moreover, we describe our participatory approach for using user-based studies with students and teachers to iteratively refine our application. We report in this paper the result of our user-based studies with teachers and students.

11.2 Literature Review

11.2.1 Augmented Reality as Emerging Technology for Educational Settings

For several years now, augmented reality has been considered a technology that will play an important role in educational settings. In 2005 and 2006, Johnson et al. (2005, 2006) forecasted augmented reality as one of six emerging technologies that would most likely enter mainstream use in educational settings by 2009–2011. However, based on the number of publications from 2004 to 2010, augmented reality did not have the same developments as pronounced as those of other emerging technologies (Martin et al. 2011).

In 2010, Johnson et al. (2010) forecasted the adoption of augmented reality around 2012–2013 mostly because of recent advances in handheld devices. Handheld devices such as smart phones are already equipped with cameras and large screens for integrating some virtual data onto the real-world scene. Other sensors such as the GPS sensor working with the gyroscope or compass can identify the phone's location and orientation, thereby displaying relevant content to user's view.

In succeeding years, Johnson et al. (2012) predict augmented reality as an emerging technology that will impact K-12 settings by 2016–2017. Although there are several exemplifying augmented reality experiences for educational settings

(Dunleavy and Dede 2014), most developments are outside of pre-college or K-12 settings such as universities and museums (Johnson et al. 2012).

11.2.2 Mobile Augmented Reality in Learning

Augmented reality was first conceptualized with the use of head-mounted displays connected to computers. However, advances in handheld devices and network technologies enabled the use of augmented reality in the mobile design space. Mobile augmented reality has become a practical technology for ubiquitous learning. Using mobile augmented reality, students can explore the real world overlayed with virtual information that would otherwise be not apparent in the natural environment.

11.2.2.1 Defining Mobile Augmented Reality

Researchers have offered several working definitions of mobile augmented reality ranging from a broad perspective to more specific ones. In a broad sense, Dunleavy and Dede (2014) identify two kinds of mobile augmented reality, namely *location-aware augmented reality* and *vision-based augmented reality*. They make this distinction based on the sensing or tracking method applied to deliver relevant content. Similarly, Zhou et al. (2008) have made such distinction between sensor-based tracking and vision-based tracking.

Location-aware augmented reality uses the GPS module of a mobile device to identify the current location of the user, whereas vision-based augmented reality uses the camera of a mobile device to identify markers (fiducial markers, pre-trained objects) in the real world using computer vision algorithms. After identifying the location or the marker, augmented reality presents digital media such as text, images, videos, and/or 3-D models that are relevant to the current location and context.

Dunleavy and Dede's definition of location-aware augmented reality and vision-based augmented reality emphasizes the characteristics of augmented reality as a context-aware system that delivers relevant content to the student. Specht et al. (2011) qualify this by defining mobile augmented reality as a specialization of context-aware systems wherein there is close synchronization of existing human senses and perception with the digital information. Their definition emphasized on the purpose of augmented reality to enhance a person's primary senses by "perceptually embedding the information into the enhanced presentation of the world."

Specht et al. (2011) identify synchronization of the real information perceived by the primary senses (vision, aural, and tactile) with the virtual information corresponding to these primary senses as key to definition of mobile augmented reality. This concept of synchronization is an extension of the previous definitions of augmented reality for visual displays. Azuma (1997) narrows this synchronization down to the sense of sight by defining augmented reality to be when "3-D virtual objects are integrated into a 3-D real environment in real time."

To be considered augmented reality, a visual display system must fulfill three requirements: a combination of real and virtual world, 3-D registration of real and virtual objects, and real-time interaction. This definition is useful for researchers in the field of computer graphics and human–computer interfaces in classifying user interfaces. However, Wu et al. (2012) recommend that educators and designers should adapt a view of augmented reality that is not confined to specific technologies. Nowadays, augmented reality can be implemented using various technologies beyond desktop computers and head-mounted displays.

All the key ideas of context-aware, enhancing perception, and synchronization or integration are embodied in augmented reality X-ray. First, virtual information is automatically contextualized to a target real object. For example, virtual roots are displayed below a tree. Second, this interaction pattern enhances the sense of sight by allowing the superhuman ability of seeing through physical barriers such as the ground. Third, it requires 3-D registration of a virtual object onto a real-world scene. Tracking must be employed to consistently position the virtual roots at all viewing angles.

11.2.2.2 Situated Learning Using Mobile Augmented Reality

Supporting situated learning is often cited as one of the key advantages of augmented reality to learning (Dunleavy and Dede 2014; Specht et al. 2011; Wu et al. 2012). Situated learning theory explains that learning takes place through the process called *legitimate peripheral participation*. Legitimate peripheral participation happens when a student increases his participation in a community of experts by interacting with their peers, experts, environments, and artifacts. One example is language acquisition wherein we start as infants uttering our first words and become sophisticated speakers with thousands of vocabulary and intuitive grasp of grammar. This learning takes place as we interact with our parents, teachers, friends in various situations and environments.

Mobile augmented reality, as a ubiquitous context-aware technology, supports situated learning because it enables students to collaborate with other people and interact with the environment. For example, Klopfer and Squire (2008) developed a mobile augmented reality game called "Environmental Detectives" wherein students pretend to be environmental scientists investigating the spread of a toxin in their campus groundwater. In this research, the mobile game supported collaboration by requiring the students to work in groups in gathering and processing information. Moreover, the mobile game supported the interaction with the environment by navigating the students and virtually drilling wells to get sample groundwater.

Another example of supporting situated learning is the "EcoMOBILE" project (Kamarainen et al. 2013). EcoMOBILE uses mobile augmented reality to navigate students around a pond and overlay relevant virtual information onto the real pond environment. In this study, teachers report that the mobile augmented reality application promoted student interaction with each other, and the pond ecosystem. In both Environmental Detectives and EcoMOBILE, students are learning how to solve problems situated in a real environment. Compared to traditional classroom instruction, learning with mobile augmented reality enables students to apply their knowledge in real-world contexts more easily (Dunleavy and Dede 2014).

Developing augmented reality X-ray directly contributes to situated learning with mobile augmented reality. Using augmented reality X-ray, students have a new way of interacting with their environment: virtually exploring interiors of real artifacts found in the real environment.

11.2.3 Research on Handheld Augmented Reality X-Ray

Mobile augmented reality affords many interaction patterns that are useful for education. Among such interaction patterns is X-ray vision wherein students can learn from exploring a real-world environment by peering into underlying structures (Specht et al. 2011). Livingston et al. (2013) define augmented reality X-ray vision as "the ability to virtually 'see through' one surface to what in reality is hidden from view." Augmented reality X-ray is rendering virtual objects onto to the real world such that the virtual object is perceived to be behind or inside the real-world object that is occluding it. Several techniques have been offered to provide an illusion of looking through objects with special care into the depth perception of users (Livingston et al. 2013). Current research studies occlusion cues to suggest depth for users when using augmented reality X-ray in handheld devices.

11.2.3.1 State-of-the-Art Implementation of Augmented Reality X-Ray

Sandor et al. (2010) introduced a "melting metaphor" to handheld augmented reality to reveal occluded points of interest in an outdoor environment. Their application gives the user an illusion of melting buildings to reveal what is hidden behind it. In a different study, Sandor et al. (2010) compared legibility when using edge-overlay X-ray to when using saliency-based X-ray. Edge-overlay X-ray uses edges found in the real-world scene as occlusion cues. In this research, we used edge-overlay augmented reality X-ray and we discussed our implementation in Sect. 4.

Instead of edges, three salient features, hue, luminosity, and motion, can be used to create "saliency maps" as described by the visual saliency model of Itti et al. (1998). This saliency map is used to decide which occluding objects in real-world scene should be kept in the real-world scene. Overall, there is no statistically significant difference in legibility between edge-overlay and saliency-based X-rays. Edge-overlay is better for scenes with high brightness and high edge surfaces. Saliency-based is better for scenes with medium-to-low brightness (Sandor et al. 2010). Aside edges and salient features, Zollmann et al. (2010) demonstrated the use of textures found in the environment as occlusion cues. In cases wherein edges, salient features and textures cannot be found in the scene, they recommend the use of synthetic details for compensation. Zollmann, et al. did not compare their X-ray technique with other augmented reality X-ray implementations.

11.2.3.2 Depth Perception in Handheld Augmented Reality X-Ray

There are very few user-based research works investigating perceptual issues such as depth perception in mobile augmented reality using handheld devices. Dey et al. (2012) were first to investigate depth perception in augmented reality X-ray with mobile devices: iPhones and iPads. They report that when using the current state-of-the-art techniques for mobile augmented reality, users underestimate distance for medium-field (beyond arms-reach to 30 m) and far-field (beyond 30 m) distances.

Dey et al. (2012) did not conduct a study for near-field (within arms-reach) distances because the target of their work is pedestrian applications such as augmented reality browsers for navigation (Dey et al. 2011). Based on empirical research (Dey et al. 2012), there was no significant difference in depth perception at varying screen resolutions. Users underestimated distances of virtual objects from themselves more on the bigger device (iPad) compared to the smaller device (iPhone). However, using the iPad allowed for better estimation of distances between two virtual objects. One of their most interesting finding is that both the tracking method and edge-based X-ray does not influence depth perception in outdoor locations.

11.2.4 Development Models for Augmented Reality Learning Applications

For developing augmented reality systems, Livingston (2005) recommends a two-step approach in solving human factors issues such as depth perception in augmented reality X-ray. First, he recommends to conduct limited perceptual tests that use only the well-designed part of the user interface. This is to ensure that there are no perceptual issues that will hinder the users to perform higher-level tasks with the interface. Second, researchers may proceed with comparing user performance on higher-level tasks using the interface against traditional methods for solving these tasks.

When researchers attempt to skip the first step, they risk testing an interface that may have perceptual and usability issues. Gabbard and Swan (2008) support this by recommending a usability engineering model for augmented reality. They proposed the iterative use of user-based studies to gain insights for the design. Their model adapts a user-centered design by iteratively refining the interface using feedback from:

A. *User task analysis*—Requirements of the user tasks are gathered and understood.
B. *Expert evaluation*—Experts examine paper mock-ups and prototypes based on design guidelines.
C. *User-based studies*—Users are observed as they perform tasks with the interface.

In the field of educational technology, researchers need to design both the augmented reality interface and the educational experience. To develop a mobile augmented reality educational game, Klopfer and Squire (2008) adapted a development process that integrates principles from rapid prototyping, learner-centered software, and game design methodologies. Their development had six phases, namely

2.4.1. *Brainstorming*—They conceptualized novel educational software platform using mobile augmented reality.
2.4.2. *Design of first instantiation*—They designed an exemplifying mobile augmented reality game by envisioning user scenarios.
2.4.3. *Development of first instantiation*—They built a "quick and dirty" rapid prototype of the exemplifying augmented reality game.
2.4.4. *Field trials of first instantiation elements*—They tested novel elements (e.g., GPS navigation software, concept of augmented reality, and basic game functions) with students and teachers.
2.4.5. *Classroom implementation of first instantiation*—They tested the performance of the whole first instantiation by asking the teachers to use it in class.
2.4.6. *Platform design for creating next instantiations*—Based on the knowledge from 1 to 5, they created a toolkit to create similar educational mobile augmented reality games for different learning scenarios.

Consistent with the work of Klopfer and Squire (2008), Dunleavy and Dede (2014) recommended the use of *design-based research* approach to study the feasibility of applying mobile augmented reality in K-12. Design-based research is a combination of methods that refines educational applications by testing them based on principles from earlier research. Aside from insights obtained from learning theories and video game design principles, design-based research uses results of field testing individual elements and entirety of the application with the target users. This type of formative research iteratively improves a mobile augmented reality application similar to user-based studies explained by Gabbard and Swan (2008).

11.3 Method

The focus of this research is to conduct user-based studies to test the applicability of state-of-the-art augmented reality X-ray to the needs of teachers and students. First, we developed a prototype of augmented reality X-ray on a tablet computer. Then, we conducted focused group discussions (FGDs) and interviews, and experiments with elementary and high school students.

The FGD was used to brainstorm about the use of mobile augmented reality in general and generate general user requirements for a first prototype. Then, we developed a demonstration of edge-based augmented reality X-ray as the first prototype. Lastly, we used this prototype for studying the student's perception when using augmented reality X-ray.

Figure 11.1 summarizes our activities in a compact model that illustrates the feedback that informs the design and development of our mobile augmented reality application. This participatory design involves our stakeholders such as teachers and students early in the design to ensure that the application suits their needs.

11.3.1 Focused Group Discussions

Two exploratory FGDs were conducted with two schools in the Philippines. The first FGD was conducted with the school principal and a school administrator from Spring Christian School. The second FGD was conducted with four teachers and 2 parents from Spurgeon School. The goals of the FGDs were to identify gaps in

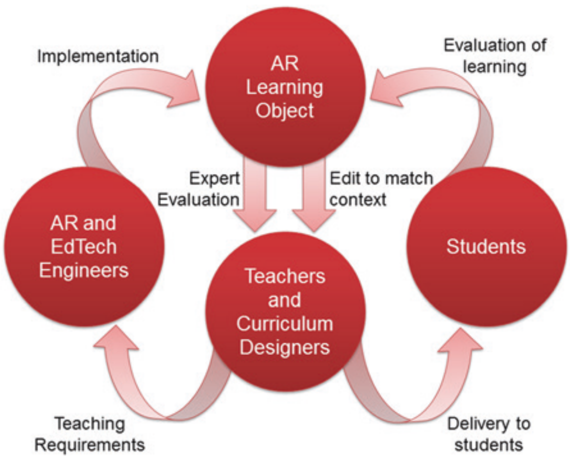

Fig. 11.1 Participatory design of augmented reality learning object (Santos et al. 2013b). This model summarizes the activities in the development of mobile augmented reality applications for educational settings

learning that could be addressed by augmented reality. Moreover, we discussed possible foreseen difficulties in classroom implementation, and their willingness to adapt augmented reality in the classroom.

As a starting point of discussion, we selected state-of-the-art prototypes designed for classroom use and sketched the application to the teachers. The applications we chose cover several topics, namely butterfly life cycle (Tarng and Ou 2012), collision in physics (Li et al. 2011), human internal anatomy (Blum et al. 2012), playing the guitar (Motokawa and Saito 2006), and magnetism (Matsutomo et al. 2012). Then, we asked the teachers open-ended questions. (e.g., What concerns come to mind when these applications presented to you? Why do you think these are useful/not useful? Why is this suitable/not suitable for this particular topic? and If you were to use this application, what would you like to remove/add?) The FGDs were conducted using colloquial language, “Taglish,” which refers to a combination of Filipino and English. This is the more natural conversation style in the Philippines. Each FGD lasted around one hour.

In both FGDs, the participants expressed interest in using the several augmented reality prototypes presented to them. The school administrators said that they will be adapting tablet computers as part of their computer laboratory soon. They see mobile augmented reality as an advantage because they can justify more concretely the need to purchase tablet computers. They also believe that one of the parent’s main concerns in choosing schools is the available facilities. Particularly, the presence of hardware and software designed for learning. As such, it is very important to keep up with the trend in technology for schools.

The teachers hesitate to use mobile augmented reality because they are not confident with using hardware and software for their class. They believe that only teachers in computer-related subjects are qualified to use augmented reality for teaching. However, they are very eager to learn the technology if curriculum designers can design lesson plans and train them to use augmented reality for their classes. Yearly, they upgrade their skills through seminars, so they recommend coursing the augmented reality trainings through these seminars.

11.3.2 Implementation of the First Prototype

The AR X-ray was implemented entirely on iPad 2 tablets with dual core Apple A5, 512-MB RAM, and 32-GB memories. We used vision-based tracking using the ARToolkit (Kato and Billinghurst 1999) which uses fiducial markers. The target real object is a cube (side = 60 cm) with print on the faces as shown in Fig. 11.2 “target object.” The virtual 3-D models displayed in the interior are cultural artifacts as shown in Fig. 11.2 “virtual object.” We used edge-based X-ray as proposed by Sandor et al. (2010). We obtain the image of the target object using the back camera of the iPad with a resolution of 480 × 640 pixel. Then, we use the

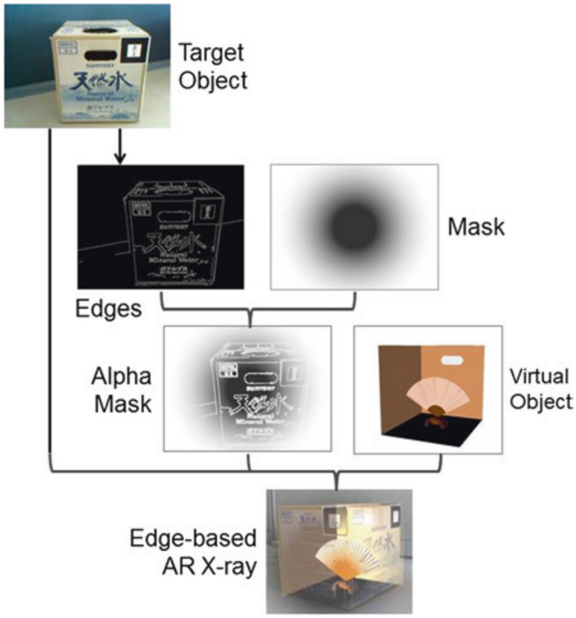

Fig. 11.2 Implementation of edge-based augmented reality

Canny edge detector to extract edges from the target object and the scene. These edges are multiplied to a radial mask to create an alpha mask. Based on the alpha mask, we blend the target object and the virtual object. The output of this blending is overlayed onto the target object.

11.3.3 Constructs for User-Based Studies

We identified three important constructs to augmented reality X-ray, namely depth perception, legibility, and realism. After identifying the important constructs, we conducted user-based studies (discussed in Sect. 5) to evaluate our augmented reality X-ray prototype (discussed in Sect. 4).

We conducted user-based studies to evaluate the augmented reality X-ray prototype. Based on the literature (Livingston et al. 2013), one of the important aspects of augmented reality X-ray is *depth perception*. Although Dey et al. (2012) have shown that augmented reality X-ray on an iPad device does not influence depth perception, their work is limited to medium-field and far-field distances. They did not test for the case of near-field or within arms-reach distances.

The second construct is *legibility* of the virtual object. Edge-based X-ray uses edges on the target object to convey depth as occlusion cues to the user. As such, the technique itself makes the virtual object less legible than in standard augmented reality. Sandor et al. (2010) have confirmed in a test that too many edges have a negative effect to legibility.

The last construct that we wanted to study is the feeling of *realism* in students. The teacher's have noted that aside from the issue of legibility, another aspect that may be confusing for students is the concept of augmented reality itself. They are interested to know how real the experience feels for the students.

11.4 The System

Figure 11.3 shows a user study participant using the augmented reality X-ray prototype. We implemented the system entirely on the iPad. It uses fiducial markers placed on the target object for tracking. In the user-based evaluations, we asked the users to explore what is inside the box using the application.

The prototype implements edge-based X-ray as described in Sect. 3.2. Figure 11.4a shows the real-world scene. Figure 11.4b shows the edges detected from Fig. 11.4a using the canny edge detector. The edge detection is suppressed on the sides as shown in Fig. 11.5c before applying to the final edge-based X-ray Fig. 11.4d.

Figure 11.5 shows the simple virtual overlay versus the augmented reality X-ray. Figure 11.5a shows the simple virtual overlay which does not include occlusion cues to provide an illusion of X-ray vision. This is used as the control scenario for the user-based studies describes in Sect. 5.

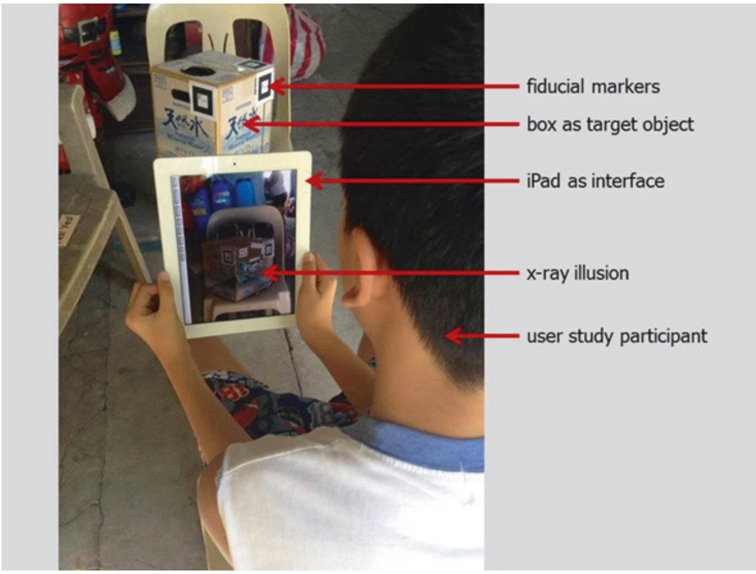

Fig. 11.3 Overview of the system

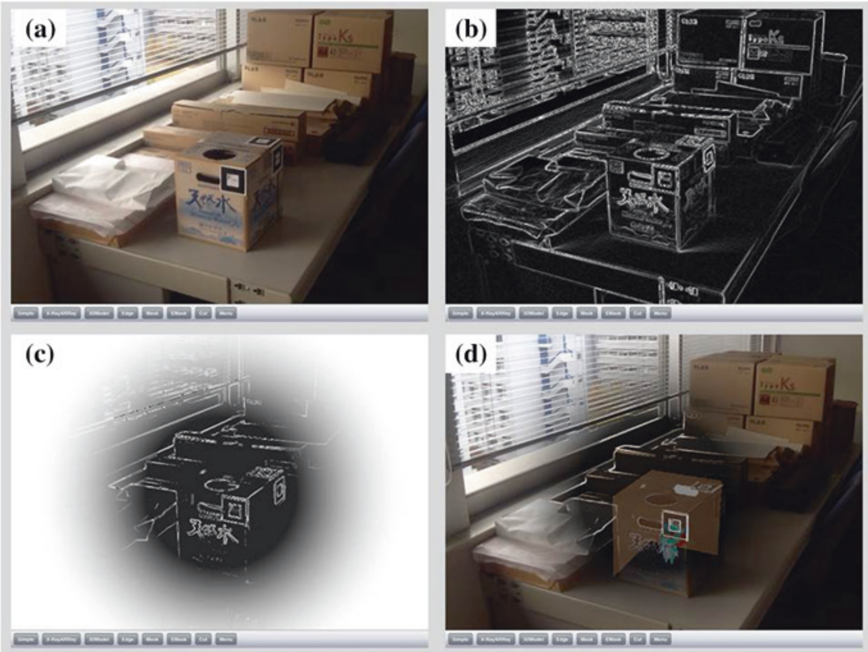

Fig. 11.4 Edge-based X-ray

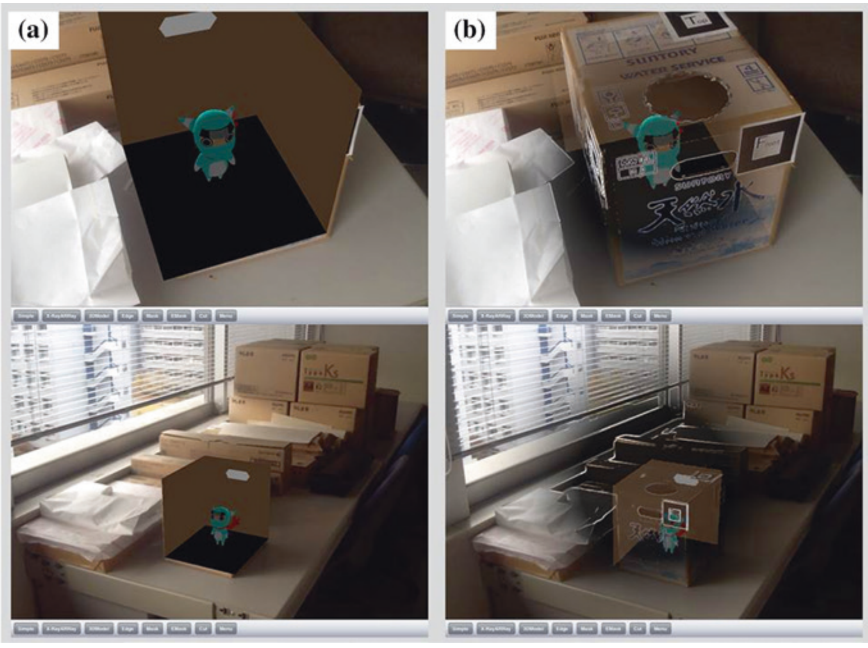

Fig. 11.5 (a) Simple virtual overlay versus (b) augmented reality X-ray

11.5 Research Design

11.5.1 Perception Evaluation

We conducted three simple evaluations of depth perception, legibility, and realism by comparing edge-based X-ray (Fig. 11.5a) and simple virtual overlay (Fig. 11.5b). The goal of the user studies is to compare whether the current implementation of edge-based X-ray influences simple virtual overlay without the edge-based occlusion cues. This simple virtual overlay is the standard augmented reality implementation. For the series of user-based evaluations, we wanted to test three hypotheses:

Hypothesis 1: The user will realize that the visualization suggests looking inside the box with augmented reality X-ray, but not for simple overlay.
Hypothesis 2: The user will perceive better legibility with simple overlay because there are no occlusion cues.
Hypothesis 3: The user will perceive more realism with augmented reality X-ray because it blends elements from the real scene onto the virtual object.

11.5.1.1 Pilot Test

The first evaluation is a pilot conducted with 12 graduate students (11 male and 1 female, aged 24–26). They were split into two groups of six with one group as control-first and the other group as experiment-first. The control-first group was shown the simple virtual overlay first. Then, they are asked to respond to 3 statements in a written questionnaire:

1. The object is inside the box.
2. The object is easy to see.
3. The object seems real.

Participants can respond to these items on a 5-point Likert scale with 1 corresponding to *Strongly Disagree* and 5 corresponding to *Strongly Agree*. After answering the short survey, they were shown the edge-based X-ray and asked to answer the questionnaire again. For the experiment-first, we reverse the viewing order such that they view edge-base X-ray first, and then they view the simple virtual overlay.

11.5.1.2 Evaluation with Students 1

The second evaluation was conducted with 23 Filipino students (9 male and 14 female, aged 5–15). The study was conducted either at the participant's home, or the home of a relative or family friend. We obtained permission from the parents to conduct this study, with a parent waiting outside the testing room. Similar to the pilot test, the participants were divided into control-first (8 participants) and experiment-first (15 participants). The students were asked to respond to six statements in an interview format:

Q1. The object is inside the box.
Q2. The object is easy to see.
Q3. The object seems real.
Q4. The object seems flat.
Q5. I can see different parts of the object clearly.
Q6. My classmates will say the object is real.

The questions were read to the students and then translated into Filipino. Students respond by picking one of five possible answers arranged in a 5-point Likert scale:

1. No! No!
2. No!
3. I don't know.
4. Yes!
5. Yes! Yes!

We included a short 5–10 min break between evaluating the control scenario and the experiment scenario. Aside from question 4, higher scores would correspond to higher perception of the construct. We inverted the response to question 4 by subtracting it from 6, thereby getting a score that is parallel with the other questions.

Lastly, we conducted a debriefing interview with the participants primarily for them to make sense of the experience. This is especially necessary because testing augmented reality systems with younger children can be a confusing experience for them. Moreover, we conducted the debriefing interview to gather the students' impression of augmented reality X-ray.

11.5.1.3 Evaluation with Students 2

The third evaluation was conducted with 47 Filipino high school students (21 male and 26 female, aged 11–16) of Spring Christian School in Muntinlupa City, Philippines. We divided the participants into two groups. The control-only group composed of 21 students viewed only the simple virtual overlay, whereas the experiment-only group composed of 26 students viewed only the edge-based X-ray. Both groups are asked to answer the same statements in *evaluation with students 1* translated into Filipino. We asked them to answer a 5-point Likert scale similar to the one in the *pilot test* translated into Filipino.

11.5.2 Teacher Interviews

We evaluated the edge-based X-ray with teachers in an interview format. We demonstrated the edge-based X-ray to twelve teachers from Spring Christian School and Spurgeon School in Makati City, Philippines. The interview flow revolved around whether or not the current implementation is appropriate and useful for their practice. We also explained two sketches of possible applications of mobile augmented reality X-ray. The first one is for looking inside the body to see the skeletal system as envisioned by Blum et al. (2012). The second one is looking inside plants to see how water is transported. Based on these two examples, teachers were asked for advantages, disadvantages, and suggestions on these proposed applications.

11.6 Evaluation and Discussion

11.6.1 Legibility Suffers in Augmented Reality X-Ray

Tables 11.1, 11.2, and 11.3 summarize the responses of the participants in the pilot study, the first and second evaluation of students, respectively. For the pilot study conducted on adults (aged 24–26), all the results were statistically significant confirming our three hypotheses. First, they confirm that augmented reality X-ray conveys that the virtual object is inside the box, which is not conveyed by the simple virtual overlay scenario. Secondly, they perceived the object to be more legible in the simple virtual overlay compared with the augmented reality X-ray. Lastly, the results show that augmented reality X-ray significantly affects the feeling of realness of the virtual object. However, we obtained different results for the user studies conducted on students. Only question 2 of the first evaluation with students confirms our second hypothesis about legibility. For rest of the items in the evaluation with students 1 and 2, results show that there are no significant differences between augmented reality X-ray and simple virtual overlay.

Our user-based studies with students do not show any significant differences (except for legibility) between viewing simple virtual overlay and edge-based X-ray. Therefore, we can proceed with next-level evaluations such as evaluation with content designed by teachers.

Table 11.1 Results of pilot test on three constructs

	Treatment	*N*	Mean	*T* value
Depth	Simple virtual overlay	12	1.2	9.7**
	Augmented reality X-ray	12	4.8	
Legibility	Simple virtual overlay	12	4.0	3.7**
	Augmented reality X-ray	12	2.7	
Realism	Simple virtual overlay	12	2.4	2.9*
	Augmented reality X-ray	12	3.0	

*:$p < 0.05$ **:$p < 0.01$

Table 11.2 Results of evaluation with students 1 on three constructs

	Treatment	*N*	Mean	Standard deviation	*T* value
Depth (Q1)	Simple virtual overlay	23	3.7	1.2	0.1180
	Augmented reality X-ray	23	3.6	1.3	
Depth (Q4)	Simple virtual overlay	23	4.1	0.9	0.5672
	Augmented reality X-ray	23	4.0	1.2	
Legibility (Q2)	Simple virtual overlay	23	4.0	1.0	2.1026*
	Augmented reality X-ray	23	3.3	1.2	
Legibility (Q5)	Simple virtual overlay	23	4.0	0.9	1.3019
	Augmented reality X-ray	23	3.6	1.3	
Realism (Q3)	Simple virtual overlay	23	3.4	1.4	1.5742
	Augmented reality X-ray	23	2.7	1.3	
Realism (Q6)	Simple virtual overlay	23	3.3	1.2	0.4762
	Augmented reality X-ray	23	3.1	1.3	

*:$p < 0.05$

Table 11.3 Results of evaluation with students 2 on three constructs

	Treatment	*N*	Mean	Standard deviation	*T* value
Depth (Q1)	Simple virtual overlay	21	3.4	1.2	0.0451
	Augmented reality X-ray	26	3.4	1.2	
Depth (Q4)	Simple virtual overlay	21	3.3	1.0	0.2054
	Augmented reality X-ray	26	3.3	1.0	
Legibility (Q2)	Simple virtual overlay	21	3.8	1.1	0.3336
	Augmented reality X-ray	26	4.0	1.1	
Legibility (Q5)	Simple virtual overlay	21	4.0	0.6	1.3216
	Augmented reality X-ray	26	3.7	1.0	
Realism (Q3)	Simple virtual overlay	21	3.1	1.2	0.6283
	Augmented reality X-ray	26	3.0	1.3	
Realism (Q6)	Simple virtual overlay	21	3.4	1.0	0.9610
	Augmented reality X-ray	26	3.7	1.0	

11.6.2 Students Describe Augmented Reality X-Ray as Addition, Placement, or the Appearance of Objects

The students under the first evaluation describe augmented reality X-ray as the addition, placement, or appearance of a virtual object (Table 11.4). All 23 students related the meaning of augmented reality X-ray with the use of the iPad because this is their first experience with augmented reality. Although 16 of the 23 participants are familiar with video games on various consoles, only 4 of them described augmented reality X-ray with keywords “3-D,” “graphics,” or “effects.”

Table 11.4 Augmented reality X-ray according to students in the first evaluation

Children's term	Meaning	*N*
"May idinadagdag"	Something is added.	7
"May nilalagay"	Something is placed.	4
"May lumalabas"	Something comes out.	12
"3-D"	Three-dimensional object	1
"graphics"	Computer graphics	1
"effects"	Special effects (e.g., movies)	2

The students did not express any confusion while using the system. However, when they reflect on their experience, they express the need to confirm whether or not the virtual object is really inside the box. In future applications of augmented reality X-ray, opening the target object may not be practical. This is the limitation of the augmented reality X-ray interaction. However, we also interpret this as arousing the curiosity in the students.

11.6.3 Augmented Reality X-Ray Provides Experiential Learning, Improved Attention, and Motivation

All twelve teachers expressed their interest in learning materials using augmented reality X-ray, and they are willing to undergo some training for using such "high-tech" materials. They believe that some topics can be illustrated more clearly to students when using augmented reality X-ray. According to the teachers, they regularly improve their skills by attending seminars or workshops. This could be a venue for learning about using augmented reality X-ray. They identified three key advantages of augmented reality X-ray for teaching:

6.3.1. Experiential learning—Eight of the teachers identified "learning by experience" as an advantage of augmented reality X-ray. Aside from learning from illustrations in books, students can be given another kind of experience that catches their attention and motivates them to learn some more.

6.3.2. Improved attention—All 12 teachers said that augmented reality X-ray will generate interest in students because their students are visual learners. Moreover, augmented reality X-ray employs some elements of video games and gadgets that their students are accustomed with.

6.3.3. Increased motivation—All 12 teachers speculate that augmented reality X-ray will motivate students by first catching their attention and then generate interest to know more. Using this novel visualization, students may be encouraged to ask more questions, or find answer for themselves in textbooks. In this scenario, augmented reality X-ray complements their practice and not replaces the need for teachers or books.

11.6.4 Augmented Reality X-Ray Should Be Cheap, Accurate, and Efficient for Classroom Use

The teachers commented three issues that need to be addressed to successfully integrate augmented reality X-ray to their practice. The first issue is the overhead costs of adapting new technology. New technology entails expenses for the school for the hardware, software, and training requirements. In the case of mobile augmented reality learning, it would require one device per child. In the Philippines, some private schools have already adapted tablet computers as replacement for some textbooks, and as part of a general computer laboratory class.

The second issue is the perception and accuracy of the virtual information. Teachers noted that augmented reality X-ray is highly subject to misinformation. The virtual abstraction provided by augmented reality X-ray may be inaccurate because of poor tracking. For example, in displaying body organs onto a student, the size of the organs and their relative positions will vary depending on the body type of the students. Inaccurate integration of the virtual objects can misinform students about their bodies.

Lastly, augmented reality X-ray can be too interactive for students and thereby consume more class time. Learning modules using augmented reality must be carefully planned to work with the time constraint allotted for the lesson. As such, we need to work with curriculum designers to create lesson plans that consider the time for teaching the necessary content.

11.7 Conclusion

We introduced “X-ray vision” as one of the key affordances of augmented reality to K-12 educational settings. This is especially useful in proposed models of mobile augmented reality learning based on situated learning. Mobile augmented reality supports situated learning because it facilitates interaction among students and with the real environment. Using augmented reality X-ray, students can interact with objects found in the real environment by virtually looking inside the target object. Recently, methods for achieving augmented reality X-ray on handheld devices have been proposed making it ideal for mobile augmented reality learning. However, there is a lack of user-based studies evaluating the state-of-the-art methods.

We described an edge-based approach to achieving the effect of “X-ray vision” and evaluated it with students and teachers. Edges from the target object are extracted and used as a mask to select which parts of the target object should be displayed as occlusion cues. We used the ARToolkit for tracking and implemented augmented reality X-ray entirely on a tablet computer. As a novel interaction for education, we investigated the student’s perception of depth, legibility, and realism in a series of user-based evaluations. Results of the evaluation with students show that except for legibility in one study, there are no significant influences of augmented reality X-ray to simple virtual overlay which is the standard case for augmented reality.

Our interviews with students did not indicate confusion in the part of the students. They were confident to describe augmented reality X-ray in their own words. The teachers were interested in using augmented reality X-ray because they think that it will give chance for students a chance to learn by experience. They believe that it will arouse interest in students and may lead them to ask more questions in class. However, key issues in monetary budget, accuracy of virtual information, and teaching time constraints must be addressed to adapt this technology.

References

Azuma, R. T. (1997). A survey of augmented reality. *Presence-Teleoperators and Virtual Environments, 6*(4), 355–385.

Bajura, M., Fuchs, H., & Ohbuchi, R. (1992). Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *ACM SIGGRAPH Computer Graphics, 26*, 203–210.

Blum, T., Kleeberger, V., Bichlmeier, C., & Navab, N. (2012). Mirracle: An augmented reality magic mirror system for anatomy education. In *IEEE Virtual Reality Workshop*, pp. 115–116.

Dey, A., Jarvis, G., Sandor, C., & Reitmayr, G. (2012). Tablet versus phone: Depth perception in handheld augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 187–196.

Dey, A., Jarvis, G., Sandor, C., Wibowo, A. K., & Mattila, V.-V. (2011). An evaluation of augmented reality X-ray vision for outdoor navigation. In *International Conference on Artificial Reality and Telexistence*, pp. 187–196.

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. *Handbook of Research on Educational Communications and Technology* (pp. 735–745). New York: Springer.

Furness, T. A. (1986). The super cockpit and its human factors challenges. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 30*, 48–52.

Gabbard, J., & Swan, J. (2008). Usability Engineering for Augmented Reality: Employing User-Based Studies to Inform Design. *IEEE Transactions on Visualization and Computer Graphics, 14*(3), 513–525.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.

Johnson, L., Adams, S., & Cummins, M. (2005). The 2005 Horizon Report.

Johnson, L., Adams, S., & Cummins, M. (2006). The 2006 Horizon Report.

Johnson, L., Adams, S., & Cummins, M. (2010). The 2010 Horizon Report.

Johnson, L., Adams, S., & Cummins, M. (2012). NMC Horizon Report: 2012 K-12 Edition.

Kamarainen, A. M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M. S., & Dede, C. (2013). EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Computers & Education*.

Kato, H., & Billinghurst, M. (1999). Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *International Workshop on Augmented Reality*, pp. 85–94.

Klopfer, E., & Squire, K. (2008). Environmental Detectives—the development of an augmented reality platform for environmental simulations. *Educational Technology Research and Development, 56*(2), 203–228.

Li, N., Chang, L., Gu, Y. X., & Duh, H.-L. (2011). Influences of AR-supported simulation on learning effectiveness in face-to-face collaborative learning for physics. In *IEEE International Conference on Advanced Learning Technologies*, pp. 320–322.

Livingston, M. (2005). Evaluating human factors in augmented reality systems. *IEEE Computer Graphics and Applications, 25*(6), 6–9.

Livingston, M. A., Dey, A., Sandor, C., & Thomas, B. H. (2013). Pursuit of "X-ray vision" for augmented reality. *Human Factors in Augmented Reality Environments* (pp. 67–107). New York: Springer.

Martin, S., Diaz, G., Sancristobal, E., Gil, R., Castro, M., & Peire, J. (2011). New technology trends in education: Seven years of forecasts and convergence. *Computers & Education, 57*(3), 1893–1906.

Matsutomo, S., Miyauchi, T., Noguchi, S., & Yamashita, H. (2012). Real-time visualization system of magnetic field utilizing augmented reality technology for education. *IEEE Transactions on Magnetics, 48*(2), 531–534.

Motokawa, Y., & Saito, H. (2006). Support system for guitar playing using augmented reality display. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 243–244.

Sandor, C., Cunningham, A., Dey, A., & Mattila, V.-V. (2010). An augmented reality X-ray system based on visual saliency. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 27–36.

Sandor, C., Dey, A., Cunningham, A., Barbier, S., Eck, U., Urquhart, D., & Rhee, S. (2010). Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. In *IEEE Virtual Reality Conference*, pp. 47–50.

Santos, M., Chen, A., Terawaki, M., Yamamoto, G., Taketomi, T., Miyazaki, J., & Kato, H. (2013). Augmented reality X-ray interaction in K-12 education: Theory, student perception and teacher evaluation. In *IEEE International Conference on Advanced Learning Technologies*, pp. 141–145.

Santos, M., Yamamoto, G., Taketomi, T., Miyazaki, J., & Kato, H. (2013). Authoring augmented reality learning experiences as learning objects. In *IEEE International Conference on Advanced Learning Technologies*, pp. 506–507.

Sotiriou, S., Anastopoulou, S., Rosenfeld, S., Aharoni, O., Hofstein, A., Bogner, F., & Hoeksema, K. (2006). Visualizing the invisible: The CONNECT approach for teaching science. In *International Conference on Advanced Learning Technologies*, pp. 1084–1086.

Specht, M., Ternier, S., & Greller, W. (2011). Mobile augmented reality for learning: a case study. *Journal of The Research Center For Educational Technology, 7*(1), 117–127.

Tarng, W., & Ou, K.-L. (2012). A study of campus butterfly ecology learning system based on augmented reality and mobile learning. In *IEEE International Conference on Wireless, Mobile and Ubiquitous Technology in Education*, pp. 62–66.

Wu, H.-K., Wen-Yu Lee, S., Chang, H.-Y., & Liang, J.-C. (2012). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*.

Zhou, F., Duh, H.-L., & Billinghurst, M. (2008). Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 193–202.

Zollmann, S., Kalkofen, D., Mendez, E., & Reitmayr, G. (2010). Image-based ghostings for single layer occlusions in augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 19–26.