Educational Technology International
2013, Vol. 14, No. 1, 109-136

Usability Study of Middle School English Digital Textbook: A Stimulated Recall Approach

Il-Hyun JO*
Ewha Womans University

Heeok HEO
Sunchon National University

Kyu Yon LIM
Ewha Womans University

Jeong-Im CHOI
Kwandong University

Jeongmin NOH
Korea Educational Research and Information Service
Korea

We conducted an empirical study with 8 middle school students in Korea to investigate the usability and usefulness of our self-designed Digital Textbook. The Stimulated Recall (SR) Method using Morae software was utilized to analyze the learners' task behaviors and mental operations while using the Digital Textbook to learn English. Collected qualitative data indicated several problems in terms of the usability and usefulness of the Digital Textbook. The findings are summarized and some implications are discussed for further revision of the Digital Textbook and validation of the SR method as a usability-usefulness test tool.

Keywords: Digital Textbook, Stimulated recall, Usability study

* Dept. of Educational Technology, College of Education, Ewha Womans University
ijo@ewha.ac.kr

Introduction

Technology-based learning for school setting is getting widely accepted innovation, offering students the opportunity to access information sources at their convenient time and place (Cotton & Gresty, 2006). However, in terms of instructional strategy, this high-tech approach only replicates traditional teaching methods without offering any innovative and desirable learning experiences (Jo, 2009). Digital learning resources that are designed and developed for self-paced learning situations may not be well blended with instructional methods in traditional classrooms. This separation between the classroom activities and the digital learning resources could have brought only a limited impact to the reform of formal educational system. One of the instructional design solutions to this problem is to introduce blended or hybrid courses, which combine traditional face-to-face with digital learning materials. Research has reported that this combination has the potential to promote active learning (Dori & Belcher, 2005).

In this regard, the experimental Digital Textbook Project, initiated and supported by the Korea Education & Research Information Service (KERIS) is aimed to develop a high impact blended learning solution to reform the nation's K-12 education system toward a technology-enhanced, learner-centered, learning environment. However, there is a paucity of instructional design guidelines for a blended learning environment that takes full advantage of the synergistic interaction among traditional instruction and digital textbook resources in the classrooms (KERIS, 2011). Thus, it is critical for us to develop an instructional design guideline before the roll-out of the Digital Textbooks to Korean educational system.

The present study has three: 1) to introduce the notion of stimulated recall (SR) as usability testing method for digital textbooks, 2) to investigate the learner experience of the Digital Textbook contents that we developed, and 3) to identify points of improvements from the classroom ecology perspective. The study findings will provide some useful design guidelines for improving the hybrid

learning environment in formal education.

As for the study method, an SR procedure was utilized. A considerable quantity of formative evaluation data relating to the existing Digital Textbooks had already been amassed, largely collected as online feedback or retrospectively in the form of questionnaire responses (Jo & Kim, 2006). While these quantitative data are of interest and provide some information on students' views about the resource, we felt that they gave little insight into the ways how it is actually being used by the end users, the students, and how the causal relationship operates between their behaviors and learning. SR, whereby the video recording is used as an 'aide-memoire' that enables participants to access their interactive decision making shortly after the event, has been used extensively in the Human-Computer Interaction (HCI) area. This technique, developed by renowned educational psychologist Benjamin Bloom (1953), aims to aid recall of concurrent cognitive activity and authentic thoughts by using an audio or video stimulus. The method has considerable potential for studies into cognitive strategies and other learning processes, particularly complex, interactive contexts characterized by novelty, uncertainty and non-deliberative behavior (Gass & Mackey, 2000; Jo & Kim, 2006; Lyle, 2003). Thus, in this study, we decided to employ an alternative method of evaluation by using an SR.

Literature Review

Usability testing of digital textbooks

Recent studies have emphasized the importance of usability testing in ICT-mediated learning environments because the usability and functionality of advanced technologies and systems may affect the successful achievement of learning goals. Usability, grounded in research on HCI and User-centered Design, is

defined as the degree to which users can perform the specified range of tasks (Crowther, et al., 2004; Lee, 1999; Mackey & Ho, 2008). According to Brinck and colleagues (2002), highly usable systems make it easy and efficient for people to achieve their goals without having excessively complicated operations (Mackey & Ho, 2008). In most cases, usability testing is very useful to gather data and information for improving the effectiveness and efficiency of the specific systems. It also provides critical impacts on educational system design and development, incorporating into the perspectives of formative evaluation (Lee, 1999). As this study focuses on the use of Digital Textbooks for achieving particular learning goals in English education, it is necessary to conduct the usability testing for improving and assuring the quality of the learning materials in a formative evaluation perspective. While the definitions and dimensions of usability testing in the previous studies vary (Badre, 2002; Crowther, et al., 2004; Lee, 1999; Nielsen, 1993), the following key attributes are applicable to the usability testing in general.

Table 1. Usability attributes

Attribute	Measurement focus
Learnability	How easy it is to understand and learn
Efficiency	How much it streamlines the learning process
Retainability	How easy it is to remember program operations
Errors	How many errors occur during work
Satisfaction	How much people enjoy using the program
Effectiveness	How well it is to accomplish the desired goals

Among the attributes explained in Table 1, this study, aiming to formatively evaluate Digital Textbooks designed for a certain learning goal, concentrates not only on how easy and efficient it is to operate as a restricted meaning of "usability", but also on its "usefulness" to help learn English as a vital condition of learning technologies for successful learning outcomes. Digital Textbooks are multimedia learning materials that provide pre-specified learning contents and activities through

ICT, and are designed for pursuing desirable learning goals. Compared to other technological products, well-designed learning materials must be easy to operate and understand (as a necessary condition), and effective to accomplish learning (as a sufficient condition) (Green, 2007; Lee, 1999). In order to test the usability of Digital Textbooks in this study, we developed five indicators, which are specified in Table 2. These indicators refer to observable moments that learners experience in learning processes via the Digital Textbooks.

Table 2. Attributes and indicators of usability of Digital Textbooks

Attribute	Indicator	Measurement focus
Learnability	Learning Flow	How well learners are immersed in successful learning
	Task Success	How well learners perform the given tasks
Effectiveness	User Errors	How many errors learners make
	System Errors	How many errors occur during work
	Efficiency	How much it streamlines the learning process

Usability can be analyzed by experts and users through various methods and techniques. Conyer (1995) explained six typical methods for usability testing: heuristic evaluation methods, pluralistic walk-throughs, formal usability inspection, empirical methods, cognitive walkthroughs, and formal design analysis (Lee, 1999). While experts evaluate the products based on usability guidelines and their past experiences with similar products, users test the assumptions of product design and development, and provide feedback to the development team (Crowther, et al., 2004). As usability testing assesses the interactions between the user and the product being tested, not on the product itself, more attention must be paid to the user testing that allows gathering the actual data of usability by the end-users and the possible influence on their learning (Koohang & Ondracek, 2005; O'Bryan et al., 2010). Also, various data collection methods can be employed for usability testing, such as questionnaire, interview/verbal report, thinking-aloud, video-recording, auto data-logging program and software support (Johnson, et al., 2007; Lee, 1999).

A questionnaire is relatively convenient to administer. An interview is more personal than a questionnaire as the interviewer works directly with the respondent. However, interviews and questionnaires may depend on respondents' internal thinking rather than external behaviors and therefore reduce the accuracy of the study findings (Polio, et al., 2006). In addition to the major techniques, thinking-aloud involves participants to talk aloud about what they are doing during an evaluation session. Even though the thinking-aloud method is useful to understand individuals' cognitive processes, it may cause cognitive overloads that interfere their effective learning performance (Branch, 2000; Jo & Kim, 2006). Video-recording is especially useful as it enables accurate recording of the participant's physical motions, facial expressions and speech (Johnson, et al., 2007), but it provides limited interpretation of internal thinking.

In order to take full advantage of the best features of some data collection methods, this study combines video-recording with interviews that delve into learners' introspection for the usability testing, which allows both observable learning behaviors and unobservable mental operations underlying the behaviors to be obtained. In particular, the SR method is employed to gain a more detailed understanding about the interaction between users and learning materials.

Stimulated recall (SR) method

SR has been increasingly used as a research tool in education in general (Egi, 2008; Lyle, 2003; Vesterinen, et al., 2010). This technique, developed by Bloom (1953), aims to aid recall of concurrent cognitive activity and authentic thoughts by using an audio or video stimulus. The basic idea underlying the SR method is that a subject may be enabled to relive an original situation with vividness and accuracy if s/he is presented with a large number of the clues or stimuli that occurred during the original situation (Bloom, 1953). The method has considerable potential for studies on cognitive strategies and other learning processes, particularly complex, interactive contexts characterized by novelty, uncertainty and non-deliberative

behavior (Lyle, 2003).

Since the Stanford Center for Research and Development in Teaching introduced SR, it has become widespread in the investigation of teachers' thinking processes and teaching actions (Vesterinen, et al., 2010). For example, Clark (1980) reviewed two studies in which SR was employed for research on teacher thinking. Video-taped materials were used to analyze teachers' thoughts, feelings and reactions in their own classrooms. Meade and McMeniman (1992) suggested SR as an alternative methodology that allows greater insights into the relationships of teacher beliefs and actions, and minimizes the possibility of superficial self-presentation on the part of the teacher. Muir and his colleagues utilized video-SR in their investigation of teachers' professional learning and its impact on their teaching practices (Muir, et al., 2010). Many researchers in this field agree that SR is more useful for examining the relationship between a person's internal thinking process and external action compared to interviewing with unreliable understanding of the relationship and observation with no access to the internal thinking (Meade & McMeniman, 1992).

However, they also criticized the shortcomings of SR in its use. Video-recording may embarrass participants and induce their unnatural behaviors. The ecology of data gathering is another threat in using SR (Vesterinen, et al., 2010). The follow-up interview using SR is conducted by a researcher's conception of critical moments in the video-recorded sessions, rather than by the interviewee's initiatives (Lyle, 2003). Various attempts to control the shortcomings of SR have been made in the field of student learning research. Gass and Mackey's detailed analyses of SR and efforts to increase its methodological robustness led to the extensive use of SR in diverse contexts of educational research (Lyle, 2003). They suggested minimizing the time delay between event and recall, and the research design of the strongest links possible between the focus of the study and the procedures for creating the recall. Evans (2009) combined the analysis of students' computer-mediated communication in second language use with stimulus recall-based introspective interviews for improving the validity and reliability of elicited data. He pointed out

the potential benefits of applying SR for understanding learners' current thinking since the learners' thinking is not exclusively anchored in accurate recall of past experiences, but is also seen as a window on their current and developing assumptions. DeWitt and Osborne (2010) conducted SR interviews with primary students about science center visits. They used two kinds of stimuli, video recordings and still photographs, to ensure accurate recall by the participants. These techniques enable participants to vividly re-engage in previous experiences and researchers to investigate participants' recall and its impact on meaning making. They also discussed the possibility of using the video and photographs as mediation for constructing meaning from individuals' experience.

Despite the on-going debate on the advantages and disadvantages of SR, it is getting more attention, especially in technology-based learning environments and e-Learning (Jo & Kim, 2006). As SR allows participants' authentic actions and cognitive processes to be analyzed through follow-up interviews, it can provide informative evidence to improve the quality of ICT use in learning environments and to investigate the effects of learning technologies for achieving desirable learning goals. For instance, Beers and his colleagues used SR interviews to obtain a more comprehensive view of the effects of learning tools (Beers, et al., 2008). As mentioned by Vesterinen, et al. (2010) and Jo, et al. (2006), the integration with ICT offers more opportunities to obtain authentic data on participants' actions and thoughts. Software such as Camtasia and Morae is very useful to gather actual information that indicates learners' interaction with computers and behavioral traces of cognitive development. This study employed the Morae software (TechSmith, 2010), which has been developed for usability testing by TechSmith Corporation. Morae facilitates the capture of users' screens during their actual use, and the recording of the users' facial and bodily expressions, and verbal utterances with externally equipped webcams and recording headsets in less obtrusive contexts. In effect, this study used screen captures and students' action records during learning processes with the material as stimuli for assessing the usability and usefulness of the Digital Textbooks in middle school English education.

Method

Participants

Ten seventh graders, two (all female) for a preliminary study and eight(six male, two female) for the main study, from a middle school located in Seoul participated in this study. The sample size of eight is larger than the so called "Magic Number 5", which asserts that a sample of 5 users suffices to reveal about 80%of all the problems that exist in the interface under evaluation (Nielsen, 2000). The participants had been exposed to the experimental Digital Textbook since the previous semester.

They were selected from the same class by the teacher. Since they had used the English Digital Textbook from the beginning of the semester, they were believed to be skillful in its utilization. Thus, the basic usability of the Digital Textbook was assumed to already have been established when the study began. However, since mostly they used the Digital Textbook in a classroom, they were not used to certain features of the Digital Textbook in self-directed study.

Material

We used the middle school English Digital Textbook† as learning material. This textbook consists of three major sections: "Listen & Speak", "Read &Do", and "Think & Write". For this study, the "Read & Do" section was selected because it includes many scaffolding features to help students understand the reading materials. The textbook also contains three different levels of activities for differentiated learning. Students were asked to look at the learning objectives at the beginning and then go to "Read & Do" section for the study. We chose a chapter

† English Digital Textbook has been developed by Ministry of Education, Science, and Technology (MEST) in Korea and Korea Education & Research Information Service (KERIS) since 2009, and distributed to the school.

that the students had not learned before and asked them to study the learning material by themselves without any instruction from the teacher.

Procedure

Preliminary study

Before the main experiment, we conducted a preliminary study to examine the experimental process and share the understanding of the SR approach among ourselves. Two students, who did not participate in the main study, participated in the test. They used one notebook computer together, in which the English Digital Textbook and the Morae software were installed. Their learning processes were videotaped both by the Morae and a camcorder. The students were asked to study a chapter of the English Digital Textbook together. They were allowed to discuss with each other in order to promote their active involvement and to vocalize their thoughts during their learning. Based on the preliminary test, we established a procedure for the main experiment, which is depicted in Figure 1.

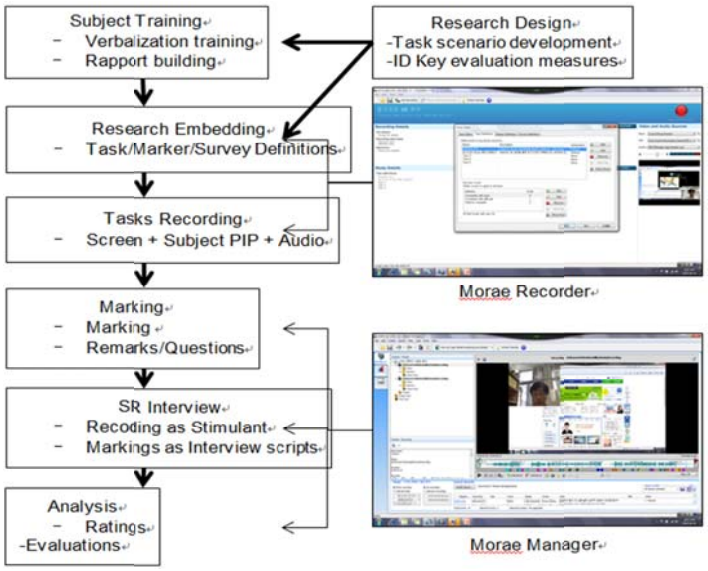


Figure 1. Procedure of the main experiment

After the preliminary test, we went through the videotape and discussed about how to conduct the SR interviews. More specifically, interview keys and marker categories were identified during the discussion, and applied to the stimuli created via Morae that were used for helping the interviewee-learners recall their mental processes during their learning with the target material (For the stimulant, see Figure 2).

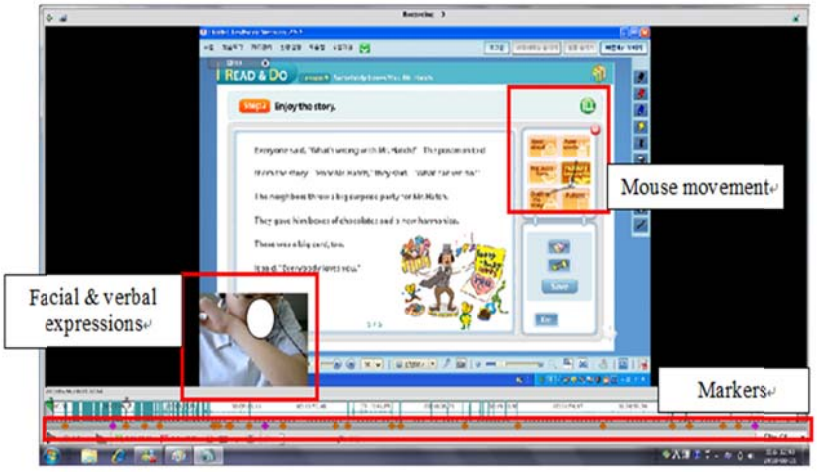


Figure 2. Screen shot of a stimulant

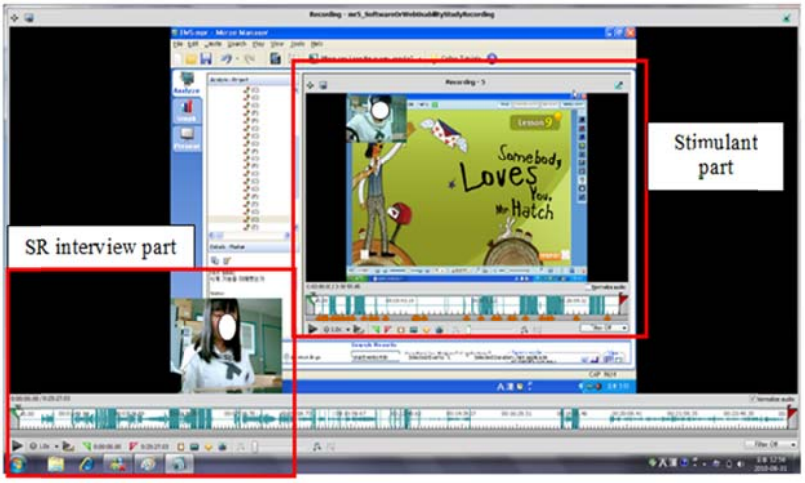


Figure 3. Screen shot of SR interview data (with embedded stimulant)

We made learners' behavioral indicators (or "markers" as Morae terminology) at the point where we needed to clarify the meaning of the facial expressions and formulated question about what we are going to ask (see lower box of Figure 2). After a discussion session among ourselves, the students joined the meeting and reviewed their videos with us. Based on the preliminary test, we decided to prepare markers to observe before the main experiment and decided to reduce the range of the Digital Textbook for the experiment. Also, since the students had problems to understand learning objectives of the chapter, we decided that the students should look at the learning objectives before the major contents for the experiment.

Main data collection

One week later, we conducted the main experiment. Eight participants used computers in a computer lab. The English Digital Textbook and the Morae software were installed in the computer. The participants studied the "Read & Do" of a chapter by themselves. While recordings of students' learning behaviors with the Digital Textbooks were made using Morae software, we also observed their learning processes and recorded them with the Morae software. Most students finished their learning in about 30 minutes, which is considered enough.

After self-learning, we reviewed each student's video and marked on the point that they wanted to ask questions. We pre-categorized the marking symbols to represent the students' behaviors, such as X-system error, U-user error, F-task failure, C-customer (student)'s behavior, T-time on task, and O-others.

Participants' use of the Digital Textbooks was recorded via the participants' audio and video in addition to the screen capture, and finally the following anecdotes were created. Figure 4 illustrates the screen viewed by the participant. To be specific, he is using the 'Chunk' function while reading the passage and the upper body of the participant is also recorded simultaneously. The participant is pointing at line 3 as he is listening and reading the passage.

Right after the initial recording, the interviewers reviewed each of the recording

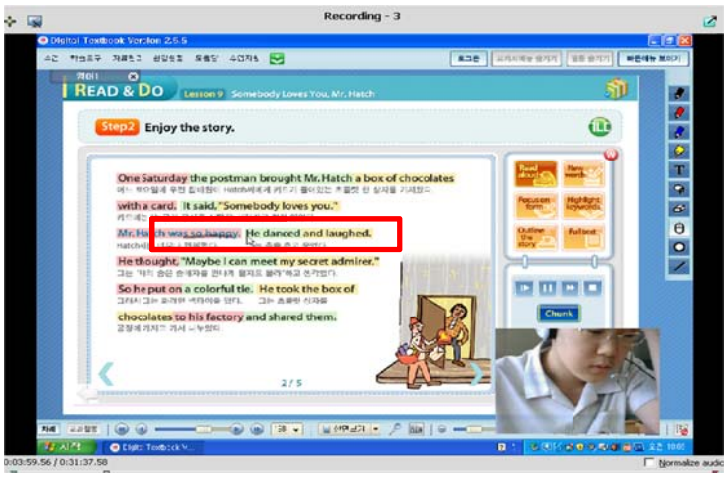


Figure 4. Screen shot of participant recording using Morae

files using Morae Manager software, and inserted "markers" throughout the timeline where there were any interesting actions or facial expressions. Pace of learning, concentrations/distractions during learning, task failures, and system error were other marking points to prepare for the SR interviews. Approximately 25 to 30 markers were generated for each of the 30-minute recordings, and the interviewers added a quick note for each marker as illustrated in the pop-up window in Figure 5.

When the recordings had been reviewed, the interviewers met the participants individually with the recordings running on the computer, and conducted interviews using the recordings and markers as stimulants.

After 3 hours, the subjects returned to participate in an SR process with us. During the SR procedure, a student viewed the captured screen and video along with the principal researcher. This researcher stopped the video at the points where the meaning of the motion needed to be clarified and asked questions to recall the student's thoughts and feelings during the learning. For example, if the researcher observed a student smiling at a certain page, s/he stopped the video and asked questions, such as "what were you trying to accomplish here?" or "what were your thoughts and feelings at this point?" or "Why did you click this button?" All

comments by the researcher and the students were simultaneously recorded in the computer by the Morae software. Each recall session lasted approximately 30 minutes.

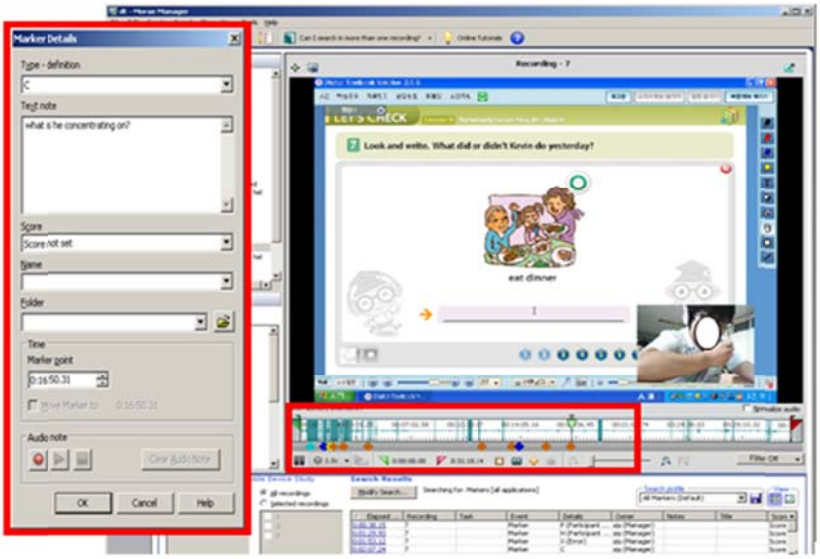


Figure 5. Screen shot of the use of the markers

Data analysis

All the data from self-learning and the SR interview were recorded by the Morae software. Morae also provided analysis tools such as simple descriptive statistics, graphics, and time checkers. Right after the data collection session, a brainstorming session was conducted to establish a shared framework of data analysis and interpretation, to secure inter-subjectivity, and to assign analysis tasks among ourselves. Following the face-to-face discussion, several follow-up online discussions were conducted when necessary. The results of the analyses and interpretations made by each researcher were sent to the principal researcher who organized them into the final analysis.

Results

Reflective notes by the researchers

Detailed reflective notes were created by each interviewer during and after the interview. These notes included (1) participants' learning processes, and (2) use of the functions and tools provided by the Digital Textbook. For example, the researcher who interviewed Participant B described the participant's learning process as follows:

> *Participant B said the chapter was quite difficult for him. He seemed to have difficulty in performing the reading task. He started reading with the listening tool turned on, and then hit the Korean translation button on page 2. Then he moved to the quiz directly, not even touching pages 3~5. He didn't read the quiz questions carefully and just clicked the next buttons to browse the other quiz questions. Finally he found that there were hint buttons for each quiz questions, but was not interested in them. After browsing the items, he went back to the reading passage to figure out the answers. But he hit quiz again in no time. He tried to type in any characters in the blank spaces, twice, in order to get the correct answers automatically. And then he copied the correct answer to the blank space, received "Great Job" feedback, and moved on to the next quiz item. The reason for this behavior was that he didn't want to waste time on something he didn't know. He just wanted to get the positive feedback and it made him feel good. Several learning tools embedded in the Digital Textbook did not work as originally designed.*

The general descriptions on the participants' learning processes provided us with a better understanding of the overall phenomenon: learning with a Digital Textbook by themselves. On the other hand, the comprehensive analysis of the interview results in terms of the participants' use of the functions and tools provided by the Digital Textbook revealed the repeated emergence of several issues.

Investigation of the usability and usefulness of functional features

Regarding research question 1, i.e., investigation of the usability and usefulness of the digital contents under the design scope of the Micro Level design, revealed valuable information that could have been overlooked. A comprehensive analysis of the interview results in terms of the participants' use of the functions and tools provided by the Digital Textbook revealed the repeated emergence of several issues. First, five out of eight participants mentioned that fill-in-the blank type activities and/or open-ended type questions were neither usable nor useful. Participant G said,

> *"The computer is not smart at all. It does not recognize my answer. I typed in a long sentence and got 'Try again' message because there was one single punctuation error. I felt really bad when this happens repeatedly."*

Related to this issue, Participant C also mentioned the low-level usability of the fill-in-the blank type activities and/or open-ended type questions.

> *"I am not good at typing in English. So I usually make some spelling errors, and it takes time to complete the quiz. It is cumbersome typing in English"*

Second, instructions displayed at the top of each page were also discussed a couple of times. Participants E and G echoed that they did not read the instructions carefully, which sometimes caused difficulty in making progress.

> *"Usually I don't read the instructions. I don't know why but it doesn't catch my eyes…. Yes, sometimes I read them when necessary, but I still don't understand what it means. I know there is a Korean/English toggle button for the instruction but the Korean instruction is not clear, either."*

Table 3. Major findings regarding the usability and usefulness of the functional features

Highly usable	**- Instructions**: All of the participants knew how to use the instructions (just simply read it, and use Korean/English toggle when they don't understand the meaning), but they don't actually use it. **- Warm-up question**: There was an open-ended question at the very beginning of the chapter in order to activate students' prior knowledge and/or other related issue. However, none of the students tried to type-in the answers. **- Non-linearity of quiz items**: When multiple questions are presented, participants read through each of them and tried to solve easier ones only. They tend to skip difficult items. **- Main idea (highlighting pen)/Text organization**: Three participants easily explored these functions, which did not actually used for their learning. **- Automatic quiz grading**: For every quiz items, participants' answers are automatically graded. However, the current system is not sensitive enough to recognize the 'almost correct' answers.	**- Show/Hide Korean translations for the reading passage**: Participants experienced no difficulty when using Korean translation while reading passages. Half of the participants mentioned that it is easy to use as well as useful. **- Listen and read**: Most of the participants said that there was no problem using listening function, and definitely helpful to listen to native speakers' reading of the passage. **- My dictionary**: A couple of participants mentioned that they use My dictionary with ease, and it is useful to keep a list of words they would like to memorize. **- Play/Speed/Volume control panel**: These are the basic functions that students used frequently for effective listening. **- Read-aloud function**: Control panel for the speed and volume of the resources were easy to use and helpful for effective learning in reading passages
Less usable	**- Feedback for the quiz**: When a participant types in his/her answer and click 'done', the screen moves on to the next page with the feedback. That is, feedback is separated from the questions, which does not follow the Contiguity Principle, causing extraneous cognitive load. **- Blank spaces for type-in answers**: Related to the aforementioned issue, participants' answers are not saved and therefore deleted when they go back to the reading passage to solve the quiz and/or when they had wrong answer. After all, participants are asked to re-type the answer from the scratch, which increase cognitive load and distractions as well.	**- Fill-in-the blank type activities and/or open-ended type questions**: Although participants had trouble in typing in English, it is necessary to learn accurate expressions, and even useful for their future learning. **- Quiz Hint**: Five of the participants struggled with using hint function. They said It was difficult to locate some of the hint buttons, and the pop-up windows were small to present all the information. On the other hand, two participants perceived that some of the hints were useful to understand the content better. **- Correct answers for the quiz**: Correct answers are presented after the participants type-in the answer. However, participants do not get the specific information regarding which part he/she made a mistake. For example, a whole sentence shows up as a correct one, rather than customized feedback.
	Less useful	**Highly useful**

Table 4. Major findings for revisions of the contents

Area	Revisions to be made
Learning aid function	For middle or low level students, additional learning aids/tools should be offered. For example, read-aloud as they read the passage, highlighting words with importance, visualizing the kernel plot structure of the passage
Dictionary function	Students unanimously appreciate the dictionary function. Enhancement of this popular feature by directly linking significant words in a passage to small dictionary applet.
Translation function	Since many students had problems because they did not understand the task instructions first of all, it is necessary to provide Korean translations as an option.
Layout design	If the passage is too long to hole in a screen, and thus should be separated into another screen, reading passage should be shown with the relevant graphical contextual material (contiguity principle should be appreciated). At the same time, full-text screen without graphics should be provided as an option for high performers.
Hints for Quiz function	Hints for quizzes should be more informative and action-based so that the students find them actually helpful for their task performance

"I don't understand what I need to do here. I didn't read the instruction. I don't click that menu because I am afraid that I might make a mistake and move to a wrong page."

Third, some of the language learning functions such as 'My dictionary', and 'Read by chunk' were supported by the participants.

"I often use 'My dictionary' because I can save the words for later use. It is very convenient. Just one click saves any words I want and creates a list for me"

"Chunk really helps me understand the meaning of a sentence when reading a long one. It shows where to stop and go."

Emerging issues regarding the participants' use of the functions and tools were incorporated into the conceptual framework of usefulness and usability that is

illustrated above. For example, instructions presented at each page were highly usable but not useful for the participants, since they usually didn't use it although they knew how. Listen and read was categorized as highly usable and highly useful as well, given that participants easily interacted with the tool and it actually helped learning. Based on these general understandings of the participants' use of Digital Textbook, the results of the SR interviews were analyzed.

Table 3 summarizes the analysis results, which enlightened us on how to improve the design of digital contents. Useful but not usable tools will definitely be redesigned. Those features that are found to be neither useful nor usable should be eliminated.

Based on the participants' behaviors and mental processes involved, several areas in the contents that require correction were identified as an answer to the research question 2. The major revisions to be made are shown in Table 4.

Reflections on the SR as research method

In terms of the methodological relevance, the SR method provided valid and rich data of user behavior and mental operations. First, the video recordings alone, even without follow-up interviews, showed the realistic learning behavior of the students and provided valuable qualitative information unobtainable with traditional quantitative research. Reviews of the video recordings to add markers enabled us to understand what the participants had experienced during the learning process. The following is a selected list of participant behaviors that may be considered as precursors to possible barriers of individual learning:

- None of the participants tried to answer the warm-up questions
- All of the students had difficulties in typing in answers to the quizzes
- When taking the quizzes, four participants directly hit 'Done'

button, copied the correct answer that was given as a feedback, and then pasted it into the black space as their answer

- Two of the participants completed the given material in 20 minutes and then browsed back and forth
- Three of the participants browsed the blue level, which was designed for advanced learners, but went back to where they were before long

Second, on top of the advantages that video recordings provided as realistic artifacts, follow-up interviews led us to better understand the underlying reasons behind the visible actions. For example, the analysis of interview transcriptions revealed the following participants' statements:

"I got a wrong answer in the quiz because I misunderstood the instruction"

"I didn't know the answer to the quiz and didn't want to spend much time on it. So I just checked the correct answer first and then copied it."

"With the teacher's help, I could've done better in the quiz."

These interview responses suggested that we need to double-check whether the instructions were described in an appropriate manner, and whether the design of providing quiz feedback was meaningful to the learners. In addition, the use of the Digital Textbook in a traditional, teacher-led classroom would have shown a different process and learning outcome. Overall, the video recordings enabled us to locate events worthy of further investigation, while the follow-up interviews deepened the understanding of the phenomenon.

On the other hand, the interview revealed the presence of concerns regarding the application of the SR method for usability testing. One of the observed behaviors

was that a couple of participants completed the material earlier than expected. The interview revealed that they did not feel like utilizing scaffolds provided to promote understanding of the reading passage.

> *"The reason why I didn't used the menus located on the right hand side was not because I didn't know that they were, but because I just didn't want to at that time."*

The response indicates that some of the participants may have not exerted full effort for learning during the experiment, despite have been asked to study the given material. This reaction was clearly evidenced by the video recordings as described above.

The other issue was the think-aloud technique. Due to the limited training for the students in verbalization during the learning task, less than expected verbal information was garnered. Verbal and facial expressions serve as cues for their introspections and cognitive strategies being operated. Lack of those cues limited the quantity and quality of the interview data.

Discussion

Regarding research question 1, investigation of the usability and usefulness of the digital contents under the design scope of the Micro Level design revealed valuable information that could have been overlooked. Emerging issues regarding the participants' use of the functions and tools were incorporated into the conceptual framework illustrated above, i.e., usefulness and usability. For example, instructions presented on each page were highly usable but not useful for the participants, since they usually didn't use them despite knowing how to. Listen and read was categorized as highly usable and highly useful as well, given that participants easily

interacted with the tool and it actually helped learning. Based on these general understandings of the participants' use of Digital Textbook, the results of SR interviews were analyzed and are summarized in Table 3.

The ecological value of the digital contents was also determined. As discussed earlier, we view the Digital Textbook as one of the many entities that reside in the classroom ecology. The experimental setting in the present study was not intended to be authentic and we, the researchers, expected some ecological problems thereof. As expected, the results indicated that the digital contents on the computers should be complemented by teachers and peers. Teacher-generated materials, individualized help for students with problems, and student-student interactions and collaborations are the examples that could enhance the usefulness of the digital contents. As we suggested at the planning stage of the project, the Digital Textbook project should be extended to the Meso Level which considers the interdependency of the ecological entities.

In terms of the methodological relevance, we found several issues. First, the SR method provided valid and rich data of user behavior and mental operations. The video recordings revealed the realistic learning behavior of the students and provided useful and rich information. Even without follow-up interviews, the video footages alone provided much valuable qualitative information that traditional quantitative research may not. Students also found their participation interesting and motivating. However, the novelty effect may have played a significant role in the students' positive reactions to the method, which renders the permanence of such motivation unknown. Second, due to the limited training for the students in verbalization during the learning task, less than expected verbal information was garnered. Verbal and facial expressions serve as cues for their introspections and cognitive strategies being operated. Lack of those cues limited the quantity and quality of the interview data.

Conclusion

The purpose of the study was three-fold: 1) to introduce the notion of stimulated recall (SR) as a usability testing method for digital textbooks, 2) to investigate the usability and usefulness of the digital contents, and 3) to identify points of improvements of the digital contents. Students' behavior captured by Morae software and metal operations by interviews with video stimulants provided rich evidence and specific information on the level of usability-usefulness of the current contents. Based on this empirically based evidence, we identified several revision points.

The findings suggest the following implications for extending to the Meso Level design model for future development projects. Students almost unanimously insisted that the Digital Textbook should be used in tandem with teachers' classroom activities, and that the digital contents could not substitute for the feedback provided by the teacher and sometimes by their peers. Some of the weaknesses of the contents revealed in the present study should be strengthened at the "outside" of the contents, which calls for additional research and development efforts regarding the Meso Level design model.

As a method of formatively evaluating the Digital Textbooks, the SR Method that we employed was found to be valid and feasible to use. If we strengthen the verbalization activities of the participants, their mental operations will be more thoroughly detected and, thereby, provide us supplementary information for instructional design activities. As discussed previously in the literature review section, the SR method was widely used to detect teacher's teaching behavior, and can be used for research on more authentic classroom situations where digital contents are blended with traditional teacher-led activities.

The present research suffered the following limitations. First, the analyses of the recordings and interview scripts were not thoroughly conducted. Student's individual characteristics and attributes were not systematically analyzed and,

therefore, some valuable implications could not be drawn. Second, the participants may have produced convincing stories about their thinking processes without knowing what really happened, and their general ways of thinking and their expectations instead of remembering their individual thinking. Third, students' level of motivation was generally low, and this could have influenced the results findings. Thus, discretion is required when to generalize the findings to the general audience. Lastly, quantitative analysis, including the number of errors, timed task efficiency, and students' perceived level of (dis)satisfaction, were not analyzed since these were beyond the scope of the present study. These limitations should be resolved in any future follow-up studies.

References

Ancona, D. G., & Caldwell, D. F. (1992). Bridging the boundary: external activity and performance in organizational teams. *Administrative Science Quarterly, 37*(4), 634-665.

Badre, A. N. (2002). *Shaping web usability: Interaction design in context.* Indianapolis, IN: Pearson Education.

Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., Gijselaers, W., & Westendorp, J. (2008). Cognitive load measurements and stimulated recall interviews for studying the effects of information and communications technology. *Educational Technology Research & Development, 56*(3), 309-328.

Bloom, B. (1954). The thought processes of students in discussion. In S. J. French (Ed), *Accent on teaching: experiments in general education* (pp. 23-46). New York: Harper.

Branch, J. L. (2000). The trouble with think-alouds: generating data using concurrent verbal protocol. CAIS 2000: dimensions of global information science. Canadian Association for Information Science Proceedings of the 28th Annual Conference.

Brinck, T., Gergle, D., & Wood, S. D. (2002). *Designing Web sites that work: Usability for the Web.* San Francisco: Morgan Kaufmann.

Clark, C. M. (1980). Choice of a model for research on teacher thinking. *Journal of Curriculum Studies, 12*(1), 41-47.

Conyer, M. (1995). User and usability testing-How it should be undertaken? *Australian Journal of Educational Technology, 11*(2), 38-51. Retrieved from http://www.ascilite.org.au/ajet/ajet11/conyer.html

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology, 37*(1), 45–54.

Crowther, M. S., Keller, C. C., & Waddoups, G. L. (2004). Improving the quality and effectiveness of computer-mediated instruction through usability

evaluations. *British Journal of Educational Technology, 35*(3), 289-303.

DeWitt, J., & Osborne, J. (2010). Recollections of exhibits: stimulated-recall interviews with primary school children about science centre visits. *International Journal of Science Education, 32*(10), 1365-1388.

Egi, T. (2008). Investigating stimulated recall as a cognitive measure: reactivity and verbal reports in SLA research methodology. *Language Awareness, 17*(3), 212-228.

Evans, M. (2009). Using stimulated recall to investigate pupils' thinking about online bilingual communication: code-switching and pronominal address in L2 French. *British Educational Research Journal, 35*(3), 469-485.

Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jo, I., & Kim. Y. (2006). An exploratory study on the validity of stimulated recall method as a usability test tool for e-Learning, *Journal of Educational Information and Media, 13*(4), 23-47.

Jo, I. (2009). Three-Level Model for the design and development of ecologically-relevant Digital Textbooks. Presented at the KELIA workshop.

Johnson, M., Ochoa, L., & Purpur, G. (2007). Virtually usable: A test of the information gardens. *Journal of Academic Librarianship, 33*(5), 593-601.

KERIS (2011). *Issue report: Digital Textbook contents development Method.* RM 2011-28.

Koohang, A., & Ondracek, J. (2005). Users' views about the usability of digital libraries. *British Journal of Educational Technology, 36*(3), 407-423.

Lee, S. H. (1999). Usability testing for developing effective interactive multimedia software: concepts, dimensions, and procedures. *Educational Technology* & *Society, 2*(2).

Lyle, J. (2003). Stimulated recall: a report on its use in naturalistic research. *British Educational Research Journal, 29*(6), 861-878.

Mackey, T. P., & Ho, J. (2008). Exploring the relationships between web usability and students' perceived learning in web-based multimedia (WBMM) tutorials.

Computers & *Education, 50*(1), 386–409.

Meade, P., & McMeniman, M. (1992). Stimulated recall - An effective methodology for examining successful teaching science. *Australian Educational Researcher, 19*(3), 1-18.

Muir, T., Beswick, K., & Williamson, J. (2010). Up, close and personal: teachers' responses to an individualized professional learning opportunity. *Asia-Pacific Journal of Teacher Education, 38*(2), 129-146.

Nielson, J. (1993). Iterative user interface design. *IEEE Computer, 26*(11), 32-41. Retrieved from http://www.useit.com/papers/iterative_design/

Nielsen. J. (2000). Why you only need to test with 5 users, Jakob Nielsen's Alterbox, Retrieved from http://www.useit.com/alterbox/20000319.html.

O'Bryan, C. A., Johnson, D. M., Shores-Ellis, K. D., Crandall, P. G., Marcy, J. A., Seideman, S. C., & Ricke, S. C. (2010). Designing an affordable usability test for e-Learning modules. *Journal of Food Science Education, 9*(1), 6-10.

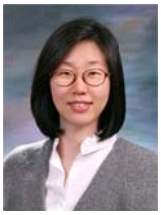
Polio, C., Gass, S., & Chapin, L. (2006). Using stimulated recall to investigate native speaker perceptions in native-nonnative speaker interaction. *Studies in Second Language Acquisition, 28,* 237–267.

TechSmith (2010). *Morae Software.* Retrieved from http://www.techsmith. com

Vesterinen, O., Toom, A., & Patrikainen, S. (2010). The stimulated recall method and ICTs in research on the reasoning of teachers. *International Journal of Research & Method in Education, 33*(2), 183-197.

Il-Hyun JO
Professor, Dept. of Educational Technology, College of Education, Ewha Womans University
Interests: educational data mining, social network analysis
E-mail: ijo@ewha.ac.kr

Heeok HEO
Professor, Dept. of Computer Education, Sunchon National University
Interests: technology based learning design, teacher education
E-mail: hoheo@sunchon.ac.kr



Kyu Yon LIM
Assistant Professor, Dept. of Educational Technology, College of Education, Ewha Womans University
Interests: instructional design model, collaborative learning
E-mail: klim@ewha.ac.kr

Jeong-Im CHOI
Professor, Dept. of Educational Technology, College of Education, Kwandong University
Interests: Instructional Design, Digital Textbook
E-mail: choij@kd.ac.kr

Jeongmin NOH
Researcher, Dept. of Digital Textbook, Korea Educational Research and Information Service
Interests: Development of digital textbook, Project management
E-mail: jamienoh@keris.or.kr

Received: March 21, 2013 / Peer review completed: April 9, 2013 / Accepted: April 29, 2013