Educational Technology International
2019, Vol. 20, No. 2, 137-168

The Effects of Information Volume and Distribution on Cognitive Load and Recall: Implications for the Design of Mobile Marker-less Augmented Reality

Taehyeong LIM*	Jiyae BONG	Ji Hei KANG	Vanessa DENNEN
Gwangju Univ. Korea	Florida State Univ. USA	Dongduk Women's Univ. Korea	Florida State Univ. USA

This study examined the effects of information volume and distribution on learners' cognitive load and recall in a mobile augmented reality (AR) environment. Information volume refers to the degree of information users are provided in a learning task, while information distribution indicates the way in which information is distributed, either in a virtual or real format. Sixteen undergraduate students participated in the study, which employed a 2 X 3 randomized block factorial design with repeated measures. Information volume and distribution were independent variables, and factors in learners' cognitive load (mental effort, perceived ease of use, and perceived task difficulty) and recall test scores were the dependent variables. Information volume had significant main effects on perceived ease of use and task difficulty, and recall test scores, while information distribution had significant main effects on perceived task difficulty and test scores. A detailed discussion and implications are provided.

Keywords : Mobile augmented reality, Cognitive load theory, Information volume, Information distribution

* Corresponding Author, Center for Teaching and Learning, Gwangju University
thlim@gwangju.ac.kr

Introduction

With the advent of mobile technologies, educational researchers have been interested in using mobile applications to promote student learning outcomes. In particular, augmented reality (AR) is one of the emerging technologies that can be easily implemented in school settings because of mobile device penetration in recent years. However, researchers in the area of mobile augmented reality studies have indicated many pedagogical and learning issues. For example, Wu et al. (2013) pointed out two major concerns: 1) a large amount of augmented information imposes cognitive overload on learners, and 2) the way in which information should be distributed between virtual and real formats. They concluded that, to overcome these issues, educational researchers and instructional designers need to be concerned with design guidelines for mobile-AR that are based on relevant learning theories and empirical research.

This study focused on two concerns mentioned above, information volume and distribution, in regard to designing learning materials in the setting of mobile augmented reality. Then Cognitive Load Theory (CLT) was employed in the study as a primary research framework in order to measure learners' cognitive burden in their learning tasks. There have been few guidelines in designing learning materials using mobile augmented reality settings. This study may illustrate cognitively optimized design principles, in certain degree, of learning materials in mobile augmented reality settings.

This study was designed to identify cognitively optimal design principles of a mobile-AR environment with respect to information volume and distribution, and various types of measurements were employed to examine learners' cognitive loads in this context. Thus, the purpose of this study was to investigate the effects of information volume and distribution on learners' cognitive loads and recall in a mobile-AR environment. Information volume refers to the degree to which users are provided with a certain volume of information in a learning task, while information distribution refers to the way in which information is distributed, either in virtual or real format.

The research questions were as follows:

1. What are the effects of different levels of information volume on learners' cognitive load and recall in a mobile-AR environment?
2. What are the effects of distributing information differently between virtual and real formats on learners' cognitive load and recall in a mobile-AR environment?
3. What are the interaction effects of information volume and distribution on learners' cognitive loads and recall in a mobile-AR environment?

Theoretical Background

Mobile augmented reality technology

Augmented Reality (AR), defined as an environment in which physical objects or content are overlaid with relevant virtual objects or content (Klopfer & Squire, 2008), has been used widely for educational purposes, including game-, place-, and problem-based learning. With the advent of mobile technology, mobile-AR has extended the advantages of AR through its unique features, including ubiquitous, collaborative, and situated learning (Wu, Lee, Chang, & Lee, 2013) in various contexts. For example, Morrison et al. (2009) introduced MapLens, a mobile-AR map application, and examined the differences between the AR map and 2D map activities. They found that AR map users were more collaborative and involved in social interactions than were 2D users. Botella et al. (2011) used a mobile-AR game to treat people with a cockroach phobia and found that the game was helpful. De Lucia, Francese, Passero, and Tortora (2012) introduced a collaborative AR system, ACCampus, which enables learners to share their contextualized information in different locations.

Mobile-AR does not refer literally to AR technology using mobile devices. According to Azuma et al. (2001), AR has three distinctive characteristics: 1) the integration of real and virtual objects in a real world; 2) a system that arranges virtual and real objects together, and 3) a system that operates interactively in real time. In addition to those characteristics, FitzGerald et al. (2013) contended that, by adding the dimension of spatial mobility, mobile AR also facilitates temporal and spatial integration of meaningful information and real-world experiences conveniently.

Mobile-AR has been established well conceptually, but has no standard technological definition. For example, with respect to devices, the technologies used most commonly are tablets and smartphones, but the use of a head-mounted display (HMD) in outdoor activities (Behzadan & Kamat, 2007) also has been referred to as mobile-AR. The technological triggers that launch relevant augmented information included marker-based AR, such as a quick response (QR) code (Pence, 2010); marker-less AR, which employs a real object as a trigger, and geo-based AR, which uses the mobile device's built-in global positioning system (GPS) to launch AR information.

Among these different types of mobile-AR technologies, this study employed specifically a mobile marker-less AR that uses a real object as a trigger to launch the relevant AR information. In addition, textual information was provided primarily in the mobile-AR environment. This particular context differs distinctly from general mobile-AR applications that include virtual objects that overlay physical objects on the same screen. Because this study included a large amount of text as the primary AR information, the virtual information cannot overlay the physical directly. Instead, an external web-link to a mobile webpage that shows the textual information overlays the physical information.

Cognitive load theory and measurement

This study employed Cognitive Load Theory (CLT) as the primary framework with which to investigate learners' cognitive processes to identify a cognitively optimal design for a mobile-AR environment. In the field of multimedia design and instruction, the CLT (Sweller, van Merriënboer, & Paas, 1998) proposed has been considered an important theoretical framework with respect to the limited capacity of human working memory. The fundamental proposition of CLT is that learners' schema construction might improve when the unnecessary cognitive load caused by inappropriate instructional design decreases. CLT assumes that a learner's cognitive load overall is determined by the sum of its three types: intrinsic, extraneous, and germane.

First, intrinsic cognitive load (ICL) is determined by the interaction between the nature of the content learned and learners' level of expertise (van Merriënboer & Ayres, 2005). Element interactivity, which is considered a main generator of ICL (Paas, Renkl, & Sweller, 2003), refers to the way the individual elements of a given task interact with others. ICL is not the same for every learner, but the level of learners' expertise plays a role in ICL's differential effects. Second, extraneous cognitive load (ECL) is associated with learning processes that are not necessary for learning directly and can be changed by instructional design and interventions (van Merriënboer & Ayres, 2005); thus, ECL is an unnecessary load attributable to the design and organization of the learning content. CLT assumes that ECL interferes with learning processes and should be reduced to the extent possible by eliminating irrelevant cognitive activities (Schnotz & Kürschner, 2007). Third, germane cognitive load (GCL) is caused by rigorous learning processes that result in schema construction and automation. While ECL affects learning processes adversely, GCL enhances them.

With respect to the measurement of cognitive load, various techniques have been suggested, but there are no absolute methods to measure it, because cognitive load is a theoretical construct that cannot be observed directly. Brunken, Plass, and Leutner (2003) summarized various cognitive load measurement methods that have been used in cognitive load research, and categorized them according to two dimensions: subjective/ objective, and indirect/direct. Subjective-Indirect methods involve a self-reported rating of mental effort. Subjective-Direct methods involve a self-reported level of stress, and the material's level of difficulty. Objective-Indirect methods involve a physiological and behavioral measure, and learning outcomes, while Objective-Direct methods involve a brain activity measure and dual-task performance.

In this study, learners' cognitive loads were measured in three ways: 1) Subjective-Indirect (a self-reported rating of mental effort); 2) Subjective-Direct (a self-reported rating of the material's level of difficulty), and 3) Objective-Indirect (learning outcomes). First, a self-reported rating of mental effort has been used frequently in cognitive load research. It was introduced first by Paas, van Merriënboer, and Adam (1994), and measures the amount of mental effort invested in understanding the learning material. Although it remains unclear how this mental effort is associated with actual cognitive load (Brunken, Plass, & Leutner, 2003), DeLeeuw and Mayer (2008) argued that a subjective rating of mental effort is a sensitive measure of ICL. Second, self-reported task difficulty also has been used often in this field. However, researchers have not reached consensus on the relation between perceived task difficulty and cognitive load. For example, Kalyuga, Chandler, and Sweller (1999) contended that a difficulty rating is related to cognitive load overall, while DeLeeuw and Mayer (2008) concluded that difficulty ratings are related to GCL. Third, learning outcomes are known to be the most common way to examine learners' cognitive loads (Brunken, Plass, & Leutner, 2003).

Methods

Participants

Sixteen undergraduate students at a flagship university in Northern Florida, US participated in this study. They were 12 female and 4 male participants whose mean age was 20.75. The group included 11 Caucasians (69%), 2 African-Americans (13%), 1 Latin-American (6%), and 2 Others (12%). To eliminate concerns about reading proficiency in the experimental task, we recruited only native English speakers. Informed consent was obtained from all individual participants included in the study.

Experimental design and variables

The experimental design was a 2 X 3 randomized block factorial design (RBF-23) with repeated measures. Two independent variables were included: 1) information volume (high vs. low) and 2) information distribution (augmented heavily vs. balanced vs. augmented lightly). The blocks in the RBF-23 consisted of six treatment conditions per participant. Repeated measures were used, so the order of presentation of the six treatments was assigned randomly for each participant. Figure 1 illustrates the experimental design and procedure in detail. The study included four dependent variables: mental effort (MEF), perceived ease of use (PEU), perceived task difficulty (PTD), and recall test score. The first three variables were categorized as cognitive load factors (Ryu & Lim, 2014), and the last was measured as a learning outcome.

		Block 1				Block 2				…	Block 6		
	Sample	Treat.	C.L. Survey	Recall Test	Break	Treat.	C.L. Survey	Recall Test	Break	…	Treat.	C.L. Survey	Recall Test
S1	O	A type (5 min)	O	O	1 min	B type (5 min)	O	O	1 min	…	F type (5 min)	O	O
S2	O	D type (5 min)	O	O	1 min	E type (5 min)	O	O	1 min	…	A type (5 min)	O	O
…	…	…	…	…	…	…	…	…	…	…	…	…	…
n	O	C type (5 min)	O	O	1 min	F type (5 min)	O	O	1 min	…	E type (5 min)	O	O

Figure 1. RBF-23 experimental design with repeated measures

The first independent variable was information volume (high vs. low). In the high volume condition, participants read and memorized text information with total word counts of 572, on average, for five minutes. In contrast, the low information volume conditions included word counts of 336, with the same time constraint. The second independent variable was information distribution (augmented heavily vs. balanced vs. augmented lightly). In the condition augmented heavily, 80% of the entire text was provided on the mobile device, while 20% was provided in print. The balanced condition provided 50% of the text information on the mobile device and the other 50% in print. The condition augmented lightly provided 20% of the text information on the mobile device and 80% in print.

The dependent variables included the cognitive load factors (CLF) and recall test score. Ryu and Lim (2016) defined CLF as multi-dimensional aspects of learners' cognitive loads in multimedia learning, and categorized them according to five factors: 1) task demand; 2) MEF; 3) PTD; 4) self-evaluation, and 5) PEU. This study employed three of them selectively. First, MEF measures the level of cognitive exertion in which the learner engages (Ryu & Lim, 2016), and increased cognitive load may play a positive role in fostering schema acquisition. Second, PEU indicates how well the learning contents are used in learning. If learners have a high level of PEU, the learning content can facilitate the learning process (Ryu & Lim, 2016). Third, PTD refers to the cognitive load that learners anticipate a given task will involve. If learners are given a complex task, their perceptions of its difficulty increase. In addition, the recall test score was another dependent variable. The assigned goal of the task was to read and memorize the text information within a given time. The learner's recall associated with the text information was regarded as the learning outcome of the task.

Selection and validity of experimental content

The selection of text content in the experiment task was one of the major concerns. The primary criteria in the selection were: 1) general information across gender and age, and 2) participants' minimum level of prior knowledge. Thus, we decided to choose movies as the experimental content with the following criteria. First, participants should have no prior knowledge of the movie. Initially, we chose titles of 20 movies released from 2009 to 2012, sorted based on the lowest number of viewers from IMDB, a web database of movies. Second, each movie title had to include a clear synopsis and two main actors/actresses, and the IMDB had to provide both the biographical information and synopsis. Third, each movie title had to have a poster showing the two main actors/actresses clearly, because we used the movie poster as a "trigger" image to launch the AR content. Through this selection process, we selected ten movie titles: *Falling Awake*; *Stolen*; *Logan*; *An Education*; *Meeting Evil*; *As Good As Dead*; *The Decoy Bride*; *Main Street*; *Passion Play*, and *Chalet Girl*.

Thereafter, a validity pilot study was conducted to identify whether our assumption of the prior knowledge of the movies selected was minimal was accurate. Ten undergraduate students (four males and six females with an average age of 22.15) participated in the validity study to measure their prior awareness of the ten movie titles and the corresponding actors/actresses. The validity scale asked their awareness of the movie titles (1-Never heard; 2-Watched a trailer; 3-Watched it on TV, and 4-Watched it in a theater), and awareness of the actors/actresses (1-Never heard; 2-Know the name, but don't know what they look like, and 3-Know what they look like). The mean of the awareness of both the movie titles and corresponding actors/actresses was calculated for each movie title. Based on the validity test results, six movie titles finally were selected and assigned randomly to each treatment condition as follows: *Stolen* (A); *As Good As Dead* (B); *Chalet Girl* (C); *Falling Awake* (D); *Logan* (E), and *The Decoy Bride* (F).

Experimental material design

Prior to designing the experimental material, the following text information for each movie title was collected: 1) movie poster image; 2) movie synopsis text, and 3) both main actors/actresses' biographic texts with portrait images. The high information volume conditions included a movie synopsis text and the two main actors/actresses' biographic texts, while the low information volume conditions included a movie synopsis text and only one actor/actress's biographic text. Each portion of the text information was divided equally. For example, in the high volume information conditions, 33% of the text information was the synopsis, 33% was a major actor's biographic text, and 33% was the other major actor's biographic text. In the low volume information condition, 50% of the text information was the synopsis and the other 50% was one major actor's biographic text.

The text information in the movie synopsis and the main actors/actresses' biographic texts was divided and then presented separately in two different formats (mobile device and print) based on the six experimental combinations shown in Figure 2. The division ratio was applied equally to each portion of the text information (synopsis and each actor/actress's biographic text). For example, in experimental condition A (high volume, augmented heavily), the entire text was provided as follows: 80% of synopsis on mobile and 20% of synopsis in print; the same percentage assignments were used the two major actor's biographic texts.

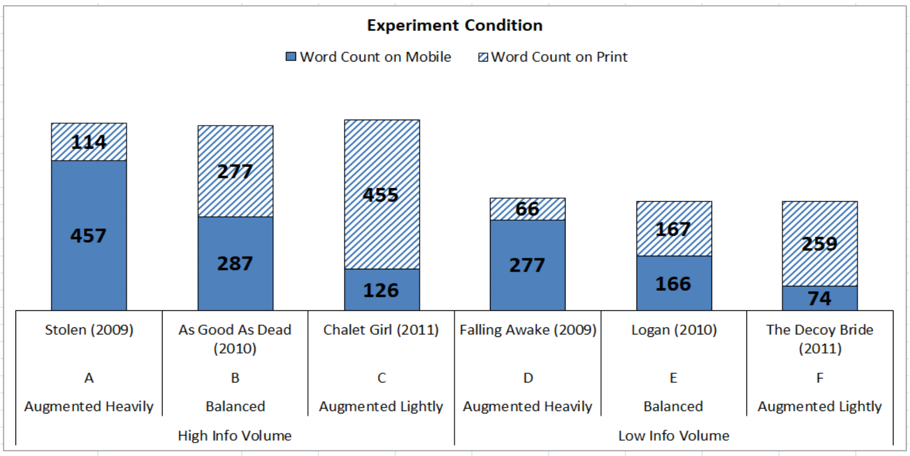

Figure 2. Design rationale of experimental materials

The text information began in print and then continued on the mobile, such that the reader was supposed to read the printed text information first and then continue to read the text on the mobile. For example, in condition D (low volume, augmented heavily), a reader was supposed to begin to read 20% of the printed synopsis, and then read the remaining 80% on the mobile, then return to the printed text to read 20% of a major actor's biographic text and then return to the mobile to read the other 80% of it. The font sizes on both print and screen were set to 10 pt.

AR tool and mobile device specification

LAYAR (2015), a mobile-AR application, was used in the study because it provides an authoring function that enables users to create AR materials easily. In addition, participants were asked to use a certain mobile device, an iPod Touch 5G, provided by the researcher.

Experimental task

The major experimental task was to read and memorize the text information about the six different movies presented both on the mobile device and on paper. The order of presentation of the movies was assigned randomly. Reading time was limited to a maximum of five minutes for each movie's content. The five-minute reading time was determined by the average reading speed associated with word counts. Rayner, Slattery, and Bélanger (2010) found that fast readers can read 330 words per minute and slow readers can read 200 words. We determined that the five-minute reading limit was sufficient even for the high volume conditions. Participants were advised to read the printed text first, and then to continue to read the text on the mobile device.

To access the AR information, participants needed to scan a movie poster as a trigger for the LAYAR application. When the AR content was launched successfully, a hovering button appeared on the mobile screen. Participants were instructed to touch the button to read the continued text information (see Figure 3).

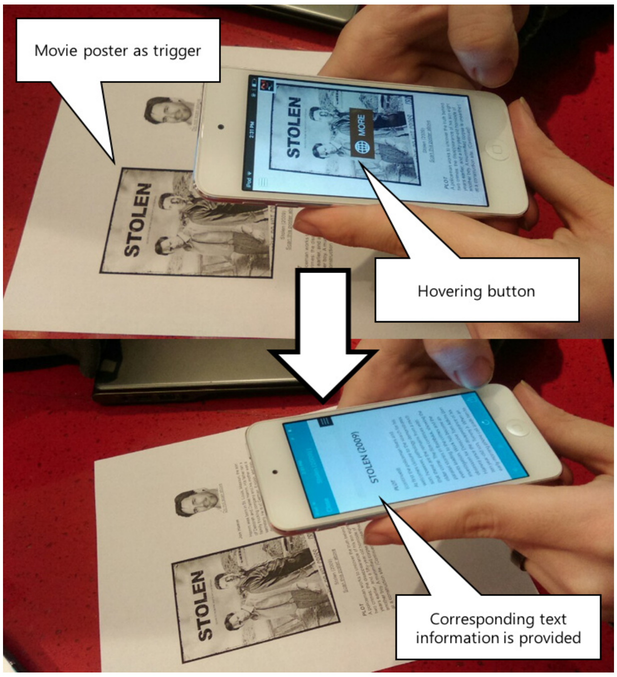

Figure 3. Users touch the hovering button to launch relevant AR information

Measures

Ryu and Lim (2016)'s CLF survey was adopted to measure cognitive load. The original five factor survey's reliability was acceptable (Cronbach's alpha = .74). Among the original five factors, three were included in this study: MEF, PEU, and PTD. Each factor included four questions measured on a 7-point Likert scale. Because the original survey was not written in English, a translation process was completed by the primary researcher, who had extensive experience with the original survey, together with a native English writing expert. The item statements also were revised slightly because the original survey focused on desktop-based multimedia learning rather than a mobile-AR context. This modified version of the three-factor cognitive load survey was administered in a paper-and-pencil format using a 5-point Likert scale.

In addition, the researcher constructed a recall test for each movie in a paper-and-pencil format. The recall test included a total of ten questions for each movie, with a score of one point per question. The high information volume conditions (A, B, and C) that included three types of textual information (synopsis, first actor/actress's biographic text, and second actor/actress's biographic text), included four questions about the synopsis, and three questions each about the first and second actor/actress's biographies: (e.g.) "In the movie, what city does Alessandra live in?" (synopsis), "what city was Jenna born in?" (actor). The low information volume conditions (D, E, and F) included two types of textual information (synopsis and first actor/actress's biography), with five questions about the synopsis and five about the first actor/actress's biography. However, condition E included four questions about the synopsis and six about the actor/actress's biography. Such a design principle for constructing the recall test was critical to create a balance between the text information participants had to memorize and the test questions they had to recall and answer (Figure 4).

	High Information Volume			Low Information Volume		
	Augmented Heavily	Balanced	Augmented Lightly	Augmented Heavily	Balanced	Augmented Lightly
	Stolen (2009)	As Good As Dead (2010)	Chalet Girl (2011)	Falling Awake (2009)	Logan (2010)	The Decoy Bride (2011)
Quiz No.	A	B	C	D	E	F
1	Synopsis(M)	Synopsis(M)	Synopsis(P)	Synopsis(M)	Synopsis(M)	Synopsis(P)
2	Synopsis(M)	Synopsis(M)	Synopsis(P)	Synopsis(M)	Synopsis(M)	Synopsis(P)
3	Synopsis(M)	Synopsis(P)	Synopsis(P)	Synopsis(M)	Synopsis(P)	Synopsis(P)
4	Synopsis(M)	Synopsis(P)	Synopsis(P)	Synopsis(M)	Synopsis(P)	Synopsis(P)
5	Actor1(M)	Actor1(M)	Actor1(M)	Synopsis(M)	Actor1(M)	Synopsis(P)
6	Actor1(M)	Actor1(M)	Actor1(P)	Actor1(M)	Actor1(M)	Actor1(M)
7	Actor1(P)	Actor1(P)	Actor1(P)	Actor1(M)	Actor1(M)	Actor1(M)
8	Actor2(M)	Actor2(M)	Actor2(M)	Actor1(M)	Actor1(P)	Actor1(P)
9	Actor2(M)	Actor2(P)	Actor2(P)	Actor1(P)	Actor1(P)	Actor1(P)
10	Actor2(P)	Actor2(P)	Actor2(P)	Actor1(P)	Actor1(P)	Actor1(P)
Note	Synopsis: questions about synopsis, Actor1: questions about actor1, Actor2: questions about actor2, (M): questions' information come from mobile screen (augmented information), (P): questions' information come from printed paper (reality information)					

Figure 4. Design rationale of recall test

Experimental procedure

The experiment was conducted individually for each participant in a small, quiet seminar room. After a participant entered the room, s/he was taught how to use the LAYAR application prior to the major experimental task. During the instruction, the participant was asked to scan a sample image on a printed paper to launch the AR information on the mobile device. Once s/he understood how to use LAYAR, a sample task exactly the same as the major task was provided as practice. S/he was asked to try to read and memorize text information about the sample movie *Titanic* within a three-minute time limit, and was told not to go forward even if s/he finished the reading within the time limit. When the time was up, the participant was asked to complete a sample recall test for the movie with no time limit, and was informed that the test score was not included in the final results. The sample practice session was conducted to help participants understand the basic structure of the major task and the recall test as well.

Thereafter, the participant was asked to complete six experimental sessions with a one-minute break between each. Each session consisted of three phases: the five-minute reading, cognitive load survey, and recall test. The procedure for each session was identical to that of the practice session described above. The only noteworthy difference was that the experimental session included a five-minute rather than a three-minute reading. Further, a cognitive load survey was included between the reading phase and the recall test. A one-minute break was provided between each session to allow participants to refresh their memories. During the break, they were asked to try to forget the movie content memorized.

Data analysis

Data were analyzed with a 2 X 3 repeated measure randomized block factorial design using SPSS v. 21.

Results

Prior knowledge of experimental content

The participants' prior knowledge of the experimental content was measured to identify any confounding variables that might affect the results of the study and to validate the assumption that the participants had minimal prior knowledge about the content. The results are provided in Table 1.

Table 1. Participants' prior knowledge of movie titles

Condition (Movie Title)	Never heard	Watched a trailer	Watched on TV	Watched in a theater
A (Stolen)	16	0	0	0
B (As Good As Dead)	16	0	0	0
C (Chalet Girl)	15	0	1	0
D (Falling Awake)	16	0	0	0
E (Logan)	16	0	0	0
F (The Decoy Bride)	16	0	0	0

Only one of the participants had watched *Chalet Girl* on TV. None of the other participants had any prior knowledge of the movie content. Based on the results, it was clear that most participants had little prior knowledge of any of the movie titles.

Descriptive statistics for dependent variables

Table 2 presents the means and standard deviations for each of the four dependent variables (MEF, PTD, PEU, and recall test score).

The participants generally demonstrated a high level of MEF (*M* = 4.61). They perceived a slightly lower level of MEF in the high volume (*M* = 4.58) than in the low volume information conditions (*M* = 4.64). The MEF in the conditions augmented heavily (*M* = 4.65) was higher than both that in the balanced condition (*M* = 4.61) and that augmented lightly (*M* = 4.58). Condition D (low information, augmented heavily) exhibited the highest MEF (*M* = 4.67), while it was lowest in condition B (high information, balanced: *M* = 4.52).

The participants also generally exhibited a high level of PEU of the material formats (*M* = 3.83). They perceived that the high information conditions (*M* = 3.63) offered less user-friendly experiences than did those in the low information conditions (*M* = 4.03). They also perceived that the conditions augmented heavily (*M* = 3.86) provided more user-friendly experiences than did those of both the balanced condition (*M* = 3.85) and that augmented lightly (*M* = 3.78). Condition E (low information, balanced) demonstrated the highest PEU (*M* = 4.13), while condition C exhibited the lowest (high information, augmented lightly: *M* = 3.56).

Overall, the participants demonstrated a moderate level of PTD of the experimental content (*M* = 2.81). Their PEU was greater in the high (*M* = 3.29) than in the low information conditions (*M* = 2.33). They also perceived that the conditions that were augmented heavily (*M* = 2.51) were less strenuous than were both the balanced condition (*M* = 2.89) and that augmented lightly (*M* = 3.04). Condition C (high information, augmented lightly) demonstrated the highest PTD (*M* = 3.60), while it was lowest in condition D (low information, augmented heavily: *M* = 2.11).

The participants also generally exhibited a moderate level of recall (*M* = 4.84 out of 10). Their recall was lower in the high (*M* = 3.10) than in the low information conditions (*M* = 6.58), and the conditions that were augmented heavily (*M* = 5.97) yielded better recall test scores than did both the balanced condition (*M* = 4.84) and that augmented lightly (*M* = 3.72). Condition D (low information condition, augmented heavily: *M* = 7.75) had the highest mean recall test score, while the recall test score was lowest in condition C (high information, augmented lightly: *M* = 2.00).

Table 2. Descriptive Statistics for Dependent Variables

Dependent Variables	Info. Distribution	Info. Volume		
		High	Low	Total
Mental Effort (MEF)	Heavily	4.63 (0.43)	4.67 (0.43)	4.65 (0.42)
	Balanced	4.52 (0.40)	4.70 (0.43)	4.61 (0.42)
	Lightly	4.59 (0.45)	4.56 (0.44)	4.58 (0.44)
	Total	4.58 (0.42)	4.64 (0.43)	4.61 (0.42)
Perceived Ease of Use (PEU)	Heavily	3.75 (0.79)	3.97 (0.54)	3.86 (0.67)
	Balanced	3.58 (0.87)	4.13 (0.64)	3.85 (0.80)
	Lightly	3.56 (0.91)	4.00 (0.63)	3.78 (0.80)
	Total	3.63 (0.84)	4.03 (0.60)	3.83 (0.75)
Perceived Task Difficulty (PTD)	Heavily	2.88 (0.81)	2.11 (0.47)	2.51 (0.76)
	Balanced	3.39 (0.89)	2.39 (0.58)	2.89 (0.90)
	Lightly	3.60 (0.78)	2.47 (0.72)	3.04 (0.94)
	Total	3.29 (0.87)	2.33 (0.62)	2.81 (0.89)
Recall Test Score	Heavily	4.31 (1.78)	7.75 (1.69)	5.97 (2.42)
	Balanced	3.00 (1.79)	6.69 (2.02)	4.84 (2.65)
	Lightly	2.00 (1.55)	5.50 (2.42)	3.72 (2.67)
	Total	3.10 (1.92)	6.58 (2.24)	4.84 (2.72)

Effects of information volume and distribution on MEF

Neither information volume [F (1, 15) = 1.60, p = .23, partial η^2 = .10], information distribution [F (2, 14) = .99, p = .40, partial η^2 = .12], nor the interaction between them [F (2, 14) = 2.57, p = .11, partial η^2 = .27] affected MEF significantly.

Effects of information volume and distribution on PEU

As shown in Figure 5, information volume had a significant main effect on PEU [$F(1, 15) = 6.19$, $p < .05$, partial $\eta^2 = .29$]. Thus, regardless of information distribution, participants in the low information conditions exhibited greater PEU associated with the experimental material than did those in the high information conditions. However, neither information distribution [$F(2, 14) = .25$, $p = .78$, partial $\eta^2 = .04$] nor the interaction between information volume and distribution [$F(2, 14) = .92$, $p = .42$, partial $\eta^2 = .12$] had a significant main effect on PEU.

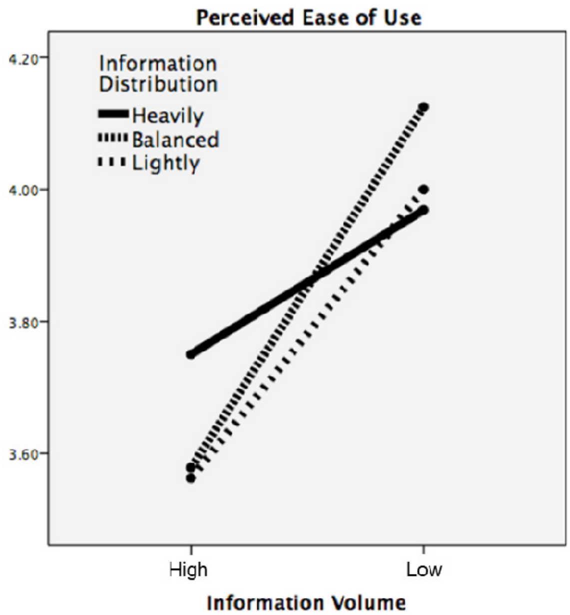

Figure 5. Information volume had a significant main effect on PEU

Effects of information volume and distribution on PTD

As Figure 6 shows, information volume had a significant main effect on PTD [$F(1, 15) = 77.64$, $p < .001$, partial $\eta^2 = .84$], which indicates that, regardless of information distribution, participants in the high information conditions perceived a greater level of task difficulty than did those in the low information conditions. In addition, information distribution had a significant main effect on PTD [F (2, 14) = 9.73, $p < .01$, partial $\eta^2 = .58$], demonstrating that, regardless of information volume, participants perceived that the tasks in the conditions that were augmented heavily were the easiest. However, there was no significant interaction between information volume and distribution on PTD [F (2, 14) = 1.15, $p = .47$, partial $\eta^2 = .14$].

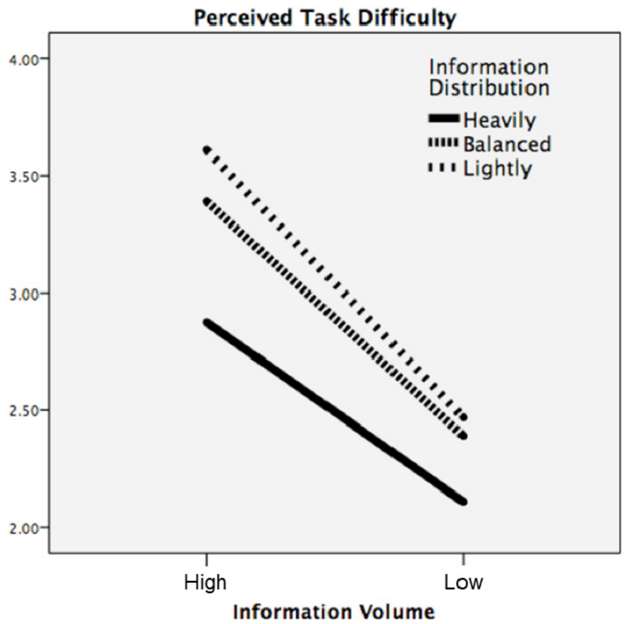

Figure 6. Both information volume and distribution had significant main effects on PTD

Effects of information volume and distribution on recall test score

As Figure 7 shows, information volume had a significant main effect on the recall test score [F (1, 15) = 157.75, $p < .001$, partial $\eta^2 = .91$]. This demonstrates that, regardless of information distribution, participants in the high information conditions had lower recall test scores than did those in the low information conditions. In addition, information distribution had a significant main effect on the recall test score [F (2, 14) = 48.98, p < .001, partial η^2 = .88]. Regardless of information volume, participants in the conditions that were augmented heavily had the highest recall test scores, compared to those in both the conditions that were balanced and augmented lightly. However, there was no significant interaction between them on the recall test score [F (2, 14) = .06, p = .938, partial η^2 = .01].

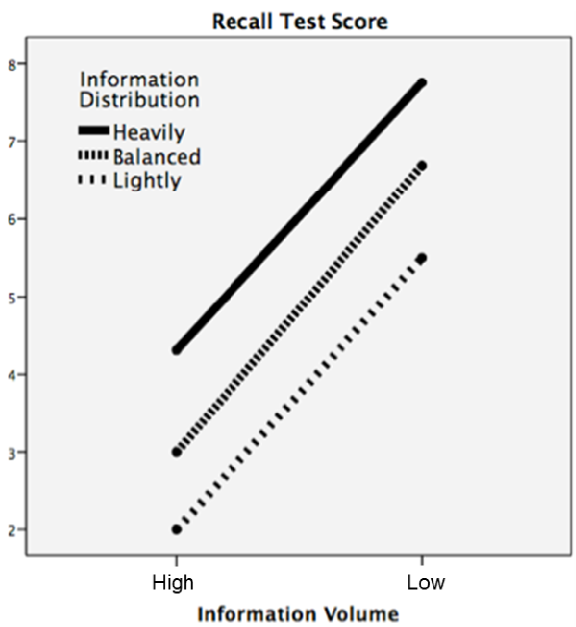

Figure 7. Both information volume and distribution had significant main effects on recall test scores

Discussion

Non-significant effects of information volume and distribution on MEF

Intrinsic cognitive load (ICL) is determined by the interaction between the nature of the learning content and learners' level of expertise. In this study, we assumed that the information volume conditions (high vs. low) would affect learners' ICL because high information volume requires learners to engage greater information processing than does low volume. Therefore, under conditions in which the learners' level of expertise is controlled, the volume of information is perhaps a key factor in differentiating learners' degree of ICL. To test this hypothesis, we measured participants' MEF using a self-report scale that was sensitive to ICL (DeLeeuw & Mayer, 2008); the more complex the materials learners were provided, the higher they rated MEF. However, the different information volume conditions had no significant effect on MEF.

There may be two reasons for this unexpected result. First, the high information condition might not have been sufficiently high for a 5-minute reading time, such that it did not pose a challenge in the high information conditions because the participants were able to read the text multiple times. Thus, participants in both the high and low information volume conditions may have perceived a similar amount of information given the period allowed. Second, the goal of the task was merely to read and memorize the content, so that the level of complexity of the tasks in both information volume conditions might have been similar. Although participants were provided with the high volume conditions, the task may still have been effectively the same as in the low volume conditions. We argue that these reasons may explain why the participants' ICLs did not differ significantly between the information volume conditions.

Significant main effect of information volume on PEU

Extraneous cognitive load (ECL) is an unnecessary cognitive burden that results from inappropriate instruction designs, regardless of the learning content. Information distribution in this study (augmented heavily vs. balanced vs. augmented lightly) was related to the layout designs of the learning materials. Thus, the study hypothesized that information distribution would impose a different level of ECL on learners, and PEU was used to test this. We assumed that PEU would differ between the information distribution conditions.

However, the study showed an unexpected result: there was a significant difference between PEU in the information volume conditions, but not in the distribution conditions. There are two possible reasons for this result. First, PEU might have failed to measure ECL even though an actual difference exists between the distribution conditions. Second, the three levels of information distribution might not have been sufficiently different to be significantly influential.

Among these possible reasons, we argue that the first is more persuasive. As explained in the next section, we found that information distribution had a significant effect on PTD as well as recall test scores. This suggests that the three levels of information distribution did differ significantly in some aspects. Thus, we favor the first reason—that PEU failed to measure ECL, even though it had an actual effect on ECL. In this respect, previous studies also have indicated that PEU, a subjective self-report measure, might not be sufficiently sensitive to measure ECL. For example, DeLeeuw and Mayer (2008) argued that objective methods (e.g., secondary task response time) were more sensitive to ECL than were subjective scales (e.g., mental effort or difficulty ratings).

In addition, we need to explicate why PEU differed significantly between the information volume conditions, but not between the information distribution conditions, as hypothesized. This might have been attributable to the meaning of the text in the measurement items. The wording in the PEU scale included "material layout," "organized," "designed," and so forth. It is possible that participants perceived that these words referred to different amount of information volume, rather than distribution. If that is the case, although these words are unrelated to information volume, participants still may have perceived that they were.

As we argued, PEU failed to measure the ECL that layout designs (information distribution types) were assumed to differentiate. However, PEU was able to detect a significant difference between information volume, even though it was not related intentionally to the measurement scale. This unexpected result indicates that there could be a confounding factor that increases ECL in certain contexts, including a device with a small-sized screen, as well as two different formats (virtual on screen and reality in print).

The confounding factor, to which we refer as a "device factor," might derive from the unique characteristics of the mobile-AR environment in the study, in which "mobile-AR" included: 1) a small-screen device and 2) two different information sources. These are device-specific contexts that differ from the traditional desktop-PC-based multimedia instruction. Specifically, two different information sources are a unique feature of AR technology that the traditional cognitive load theory approach cannot explain. We concluded that this device factor may be a unique component that contributes to ECL in a mobile-AR environment.

Significant main effects of both information volume and distribution on PTD and recall test scores

Both PTD and recall test scores had significant main effects on both information volume and distribution. With respect to PTD, we considered two major approaches. DeLeeuw and Mayer (2008) concluded that difficulty ratings are a sensitive measure of GCL. In contrast, Kalyuga et al. (1999) argued that difficulty ratings measure cognitive load overall. Between these two different arguments, the results of this study showed robust agreement with DeLeeuw and Mayer's (2008) argument; learners who perceived that task difficulty was greater tended to demonstrate less GCL for the following reasons. First, we assumed that, with respect to information volume only, one's cognitive load might be generated as illustrated in Figure 8. With this assumption, there might be no difference in ICL according to information volume, because MEF did not differ significantly. Although, theoretically, ECL is unrelated to information volume, we argued that a possible confounding factor (e.g., device) may contribute to ECL. Thus, based on the same level of ICL, if there is a possible difference in ECL, a free-up capacity may have been used to enhance GCL, which is helpful for learning outcomes that were related directly to better recall test scores in the low information condition.

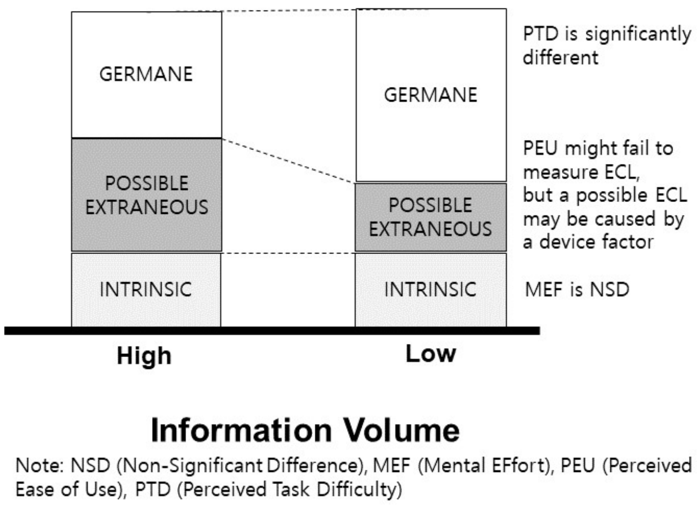

Figure 8. Assumed model of learners' GCL with respect to information volume

Second, we assumed that cognitive load might be generated with respect to information distribution only as shown in Figure 9. Information distribution theoretically is unrelated to ICL, and there also was no significant difference in MEF. In addition to the same level of ICL, we argue that there might be possible differences in ECL between the information distribution conditions, even though PEU may have failed to measure ECL. A greater free-up capacity may exist in the conditions augmented heavily, because PTD was lowest in those conditions and highest in the conditions augmented lightly, indicating that participants perceived that the conditions augmented heavily were better than were those augmented lightly with respect to information distribution types. Thus, we concluded that a greater free-up capacity exists in the conditions augmented heavily, so that a greater germane cognitive load may be generated in those conditions that results in better recall test scores.

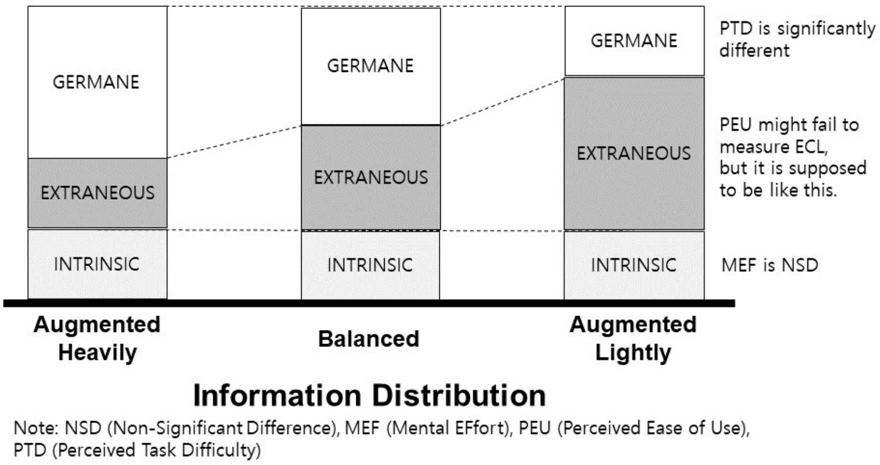

Figure 9. Assumed model of learners' GCL with respect to information distribution

Conclusions

This study examined the effects of information volume and distribution on learners' cognitive loads and recall in a mobile marker-less AR environment. With respect to various types of AR technology, a mobile marker-less AR is defined as one that: 1) uses a mobile device, and 2) a real object as a trigger. In the research design, information volume (high vs. low) and distribution (augmented heavily vs. balanced vs. augmented lightly) were included as two key independent variables. Cognitive load theory was employed to explain learners' cognitive processes that could be differentiated by the independent variables in the study. Three cognitive load factors (MEF, PEU, and PTD) and recall test scores were included as dependent variables to measure participants' cognitive loads and learning outcomes. Text information was provided for learners in the experimental task primarily in the mobile-AR environment.

The results showed that information volume had significant main effects on PEU, PTD, and recall test scores, while information distribution had significant main effects on PTD and recall test scores. Based on these results, we concluded that high information conditions are less effective in learning than are low information conditions, because one's cognitive capacity is more likely to be overloaded in the high than low information conditions. We also concluded that a large amount of text information is less suitable for processing on a device with a small screen.

In addition, with respect to information distribution, learners perceived that conditions augmented heavily made learning simpler than those augmented lightly. Conditions augmented heavily consisted of 80% of text information presented on the mobile screen and 20% in print. In this case, participants were less likely to experience cognitive overload when most of the text was provided on mobile screen. Thus, reading on a mobile screen may be more convenient for learners than reading text on paper.

These results contribute strongly to readability studies in the field of human-computer-interaction. Previous studies have shown "…no consensus on the advantages and disadvantages between paper-based reading and digital reading" (Hsieh, Kuo, & Lin, 2016, p. 435). For example, Oborne and Holton (1988) argued that there is no difference between reading on paper and on a screen, while Mangen, Walgermo, and Brønnick (2013), and Rasmusson (2015) concluded that reading on paper is better than on a screen. However, in this study, reading on a small screen was an additional factor that few studies have yet explored in detail. In addition, various confounding factors, such as age, comfort with technology, reading purpose and content, and so forth, should be considered in the context of readability studies. This study included only young participants and casual reading content (e.g., movie) in the experimental task, and these contexts should be considered when interpreting the results. Although people in general, can read better on printed paper, the young participants in this study were able to read movie content on a small screen device better.

The results of the study have the following theoretical implications. Previous studies have argued that difficulty measures are related differently to cognitive loads. For example, Kalyuga et al. (1999) argued that difficulty ratings are related to cognitive load overall, while DeLeeuw and Mayer (2008) argued that difficulty ratings are related to GCL. This study supported the latter's argument that difficulty ratings measure GCL and result in the better learning outcomes found in the recall test scores in this study. Cognitive load theory may need to be re-conceptualized in a mobile-AR context, because it differs from a desktop-based multimedia instruction environment. Mobile-AR has at least two different characteristics compared to traditional multimedia instruction. First, the screen size of the mobile device could play a critical role in interacting with the amount of text information. Second, it can present information in two different formats (virtual on screen and in the real environment) that are related to the way in which information is distributed. Thus, information distribution types (i.e., augmented highly vs. balanced vs. augmented lightly) play a vital role in affecting one's cognitive load while learning.

Moreover, the practical implications of the study indicated that conditions that are augmented heavily might be advantageous when text information is used as the primary learning material. In a mobile learning environment, it could be a disadvantage to have a large amount of information on paper vs. in a digital form. Instructional designers or mobile-AR designers should present a large amount of information in an AR format and a small amount of information in a real environment.

This study included multiple limitations and therefore, we propose further investigations of the subject. First, the distinction between the conditions with high and low information volume was determined arbitrarily. The high volume information may have not been sufficiently large to differentiate it from the low information conditions. Future studies may consider using a clear criterion to distinguish between high and low information conditions.

Second, text information largely was employed in both virtual and real space. AR technologies include various types of information, such as images, videos, audio, 3D objects, and so forth. These different formats need to be examined further with consideration of cognitive load management strategies (e.g., Modality Effect).

Third, this study was conducted in a controlled environment. Although a mobile device was used, the participants were asked not to move around the experimental site. In addition, marker-less AR is merely one type of AR technology. There also are marker-based (e.g., QR codes) or geo-location-based AR, and future studies should address these factors.

Fourth, participants' learning outcomes were measured by recall test scores, which assess only a low level of knowledge transfer. Future research should investigate learning outcomes that test comprehension, knowledge construction, or a higher level of knowledge transfer.

Fifth, movie content was employed as the primary learning content, and is not related to formal learning curricula. Instead, major learning areas, such as math and science, can be included as subject matter in future studies.

Finally, the cognitive load factor scale was developed originally based on desktop computer-based multimedia instruction, although we revised it for the AR environment. This means that there could be other confounding factors that should be considered in AR environments (e.g., device factor: screen size) which differs clearly from traditional computer-based multimedia instruction.

References

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE, 21*(6), 34-47.

Behzadan, A. H. & Kamat, V. R. (2007). Georeferenced registration of construction graphics in mobile outdoor augmented reality. *Journal of Computing in Civil Engineering, 21*(4), 247-258.

Botella, C., Breton-López, J., Quero, S., Baños, R., García-Palacios, A., Zaragoza, I., & Alcaniz, M. (2011). Treating cockroach phobia using a serious game on a mobile phone and augmented reality exposure: A single case study. *Computers in Human Behavior, 27*(1), 217-227.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

De Lucia, A., Francese, R., Passero, I., & Tortora, G. (2012). A collaborative augmented campus based on location-aware mobile technology. *International Journal of Distance Education Technologies, 10*(1), 55-73.

DeLeeuw, K. E. & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223-234.

FitzGerald, E., Ferguson, R., Adams, A., Gaved, M., Mor, Y., & Thomas, R. (2013). Augmented reality and mobile learning: The state of the art. *International Journal of Mobile and Blended Learning, 5*(4), 43–58.

Hsieh YC., Kuo CT., & Lin H. (2016). The effect of screen size of mobile devices on reading efficiency. In J. Zhou & G. Salvendy (Eds.), *Human aspects of IT for the aged population. Design for aging. ITAP 2016. Lecture notes in computer science, Vol 9754*. Springer, Cham.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371.

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.

Layar (2015) [Mobile application software]. Amsterdam, The Netherlands.

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, *58*, 61-68.

Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., Näsänen, G., & Juustila, A. (2009). Like bees around the hive: a comparative study of a mobile augmented reality map. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 1889-1898. doi:10.1145/1518701.1518991

Oborne, D. J., & Holton, D. (1988). Reading from screen versus paper: there is no difference. *International Journal of Man-machine Studies*, *28*(1), 1-9.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, *38*(1), 1-4.

Paas, F. G., van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1), 419-430.

Pence, H. E. (2010). Smartphones, smart objects, and augmented reality. *The Reference Librarian, 52*(1-2), 136-145.

Rasmusson, M. (2015). Reading Paper-Reading Screen-A Comparison of Reading Literacy in Two Different Modes. *Nordic Studies in Education*, *34*(01), 3-19.

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, *17*(6), 834-839.

Ryu, J. & Lim, T. (2014). The effects of E-book reading purposes and screen formats on reading performance and cognitive loads of college students. *Nagoya Journal of Higher Education, 14*, 197-214

Schnotz, W. & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469-508.

Sweller, J., van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296.

van Merriënboer, J. J., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development*, *53*(3), 5-13.

Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education, 62*, 41-49.

Taehyeong LIM
Assistant Professor, Center for Teaching and Learning, Gwangju University.
Interests: Mobile Learning, Social Networks, Augmented Reality
E-mail: thlim@gwangju.ac.kr

Jiyae BONG
Doctoral Candidate, Instructional Systems & Learning Technologies, College of Education, Florida State University.
Interests: Online Learning, Technology-integrated Assessment, Design Research
E-mail: jbong@my.fsu.edu

Ji Hei KANG
Assistant Professor, Library and Information Sciences, Dongduk Women's University.
Interests: School Librarians, Digital Library, Mobile Augmented Reality
E-mail: jhkang@dongduk.ac.kr

Vanessa DENNEN
Professor, Instructional Systems & Learning Technologies, College of Education, Florida State University.
Interests: Social Networks, Online Learning, Educational Technology
E-mail: vdennen@fsu.edu

Received: August 27, 2019 / Peer review completed: September. 9, 2019 / Accepted: September. 16, 2019