Examining the Effects of Different Multimedia Learning Environments on the Learning Outcomes of Second Language Learners

ARAYA RAMSIN
Chulalongkorn University Language Institute, Thailand
arayaramsin@gmail.com

HAYLEY MAYALL
Northern Illinois University, USA
hmayall@niu.edu

This quasi-experimental quantitative study examined the effects of different multimedia learning environments on the learning outcomes of second language learners. The study was designed to specifically assess Mayer’s redundancy principle by investigating whether there was a difference in performance outcomes of second language learners participating in two different multimedia learning environments: 1) images with audio only and 2) images with audio and on-screen text.

A repeated-measures analysis of covariance (ANCOVA) was conducted to evaluate change in comprehension performance from pretest to posttest, mean differences in comprehension performance between the two groups, and group differences in comprehension growth from pretest to posttest, using gender and English proficiency as control variables. Additionally, an analysis of covariance (ANCOVA) was also carried out to assess group differences in posttest scores, controlling for pretest scores, gender, and English proficiency.

Results showed no statistically significant difference in knowledge growth between participants in the two conditions and no statistically significant difference between groups in knowledge posttest scores. However,
there was a significant change from pretest to posttest for the combined sample of the learners. Additionally, a significant group × gender interaction effect was observed, where males performed better on the knowledge posttest under the Images with Audio condition while females performed better under the Images with Audio and On-Screen Text condition. Results from this study suggested that for second language learners, the presence of on-screen text, although redundant to the information contained in the audio, did not appear to impair their learning.

INTRODUCTION

The way people learn has been revolutionized through the use of multimedia for several decades (Al-Abbas, 2012; Schilling, 2009; Woo, 2009). Multimedia learning takes place when students learn from instructional content presented in more than one mode, such as words, which can be printed or spoken, and pictures, which can be static or dynamic (Issa et al., 2011; Mayer, 1997; Mayer & Moreno, 2003). Extensive research has repeatedly reported the effectiveness of multimedia learning over traditional classroom learning, in which instructions and lessons are delivered mainly through words (spoken and written). Specifically, students have been found to be more engaged and motivated in their learning across multiple domains, which often leads to a direct impact on their learning outcomes. Studies have shown this to have happened in fields such as sciences (Chuang & Liu, 2012; Moussa-Inaty & Atallah, 2012; Trevisan, Oki, & Senge, 2010), health informatics (Schilling, 2009), and language learning (Winke, Gass, & Sydorenko, 2010).

In the past few years, the focus of research has tended to shift away from evaluating the effectiveness of multimedia use in the classroom to investigating how to design multimedia learning environments that are successful and beneficial to students with various abilities and needs (Chuang & Liu, 2012; Lusk, Evans, Jeffrey, Palmer, Wikstrom, & Doolittle, 2009). Now that multimedia integration in the classroom has become a more common practice across subject areas, understanding how to utilize multimedia so that it best facilitates learners is important (Moussa-Inaty & Atallah, 2012).
PROBLEM STATEMENT

Multimedia technology plays a remarkable role in promoting effective classroom teaching and learning as teachers can use multimedia technology to transform their classes into more colorful and engaging ones (Chuang & Liu, 2012; Pun, 2013). In this age of globalization and information technology, where multimedia technology becomes more prominent and widely available to all of us (Bruce, 2014; Rana, 2013; Whitchurch, 2006), educational institutions across the globe have already started to integrate multimedia technology into education. The current audience of multimedia instruction, therefore, includes a wide range of different groups of learners, including those who are not native speakers of the language multimedia instructional materials are made in, and among them are second language learners (Rana, 2013). As an English instructor who has taught non-native English learners at universities in a foreign country for several years, the researcher has repeatedly observed that when participating in a multimedia learning environment where the language used is a language other than their mother tongue, learners tend to encounter some difficulties in comprehending the content being presented. Previous research suggests that one way to deal with this is to incorporate on-screen text, or captions, into multimedia instructional materials presented to second language learners (Chai & Erslam, 2008; Danan, 2004; Winke, Gass, & Sydorenko, 2013). However, this appears to contradict Mayer’s redundancy principle, which asserts that the visual information, or on-screen text, presented simultaneously to the pictorial and verbal information becomes redundant material, and this can result in cognitive overload (Clark & Mayer, 2008; Mayer, 2002).

THEORETICAL FRAMEWORK

The theoretical framework of this study is based on (a) the cognitive theory of multimedia learning by Richard Mayer (2002) and (b) the second language acquisition theory by Stephen Krashen (1982).

Background about Multimedia Learning

The term multimedia can be defined in a variety of ways, and it can refer to a wide range of things (Mayer, 2002). It can sometimes be used to
refer to the technologies for information display, such as computers, networks, or devices; the representational format, such as text, graphics, or animation; or the sensory modalities in perceiving information (Horz & Schnoz, 2008). According to Mayer (2002, 2014), multimedia is simply defined as the presentation of words and pictures. He further explains that words refer to material presented in verbal form, such as printed text or spoken text, and pictures refer to material presented in pictorial form, such as static graphics, including illustrations, graphs, maps, or photos, or dynamic graphics, including animation or videos (Mayer, 2014). Some examples of multimedia may include watching a podcast on a smartphone, playing a strategy game on a tablet, or watching a video on a TV screen while listening to the corresponding words, music, and sounds. Multimedia also includes any presentation in which words and pictures are integrated (Mayer, 2014). Multimedia learning, on the other hand, is defined as the learner's construction of knowledge from words and pictures (Mayer, 2014), and the process by which the learner builds mental representations from words and pictures is the focus of the cognitive theory of multimedia learning (Mayer, 2009).

Cognitive Theory of Multimedia Learning

There are many different ways to design multimedia-based instruction, and doing so is a complex challenge. One of the theories that has received wide acceptance in the world of multimedia design is the cognitive theory of multimedia learning, a theory put forward by Richard Mayer (2002). The theoretical foundation of the cognitive theory of multimedia learning draws from several cognitive theories including Paivio's dual coding theory (Paivio, 1986), Baddeley's model of working memory (Baddeley, 1992), and Sweller's cognitive load theory (Chandler & Sweller, 1991). The fundamental proposition of this theory is that students learn more deeply from words and pictures than from words alone because students have the opportunity to build meaningful connections between words and pictures (Mayer, 2002).

Mayer's cognitive theory of multimedia learning is premised on three assumptions: (1) the dual-channel assumption—humans process visual and auditory information in two different cognitive channels, (2) the limited-capacity assumption—each cognitive channel has a limited capability, and (3) the active-learning assumption—humans actively process this visual and auditory information as they learn. According to the theory, the human
information-processing system includes three memory stores called sensory memory, working memory, and long-term memory (Mayer, 2002).

The cognitive theory of multimedia learning holds that in order for meaningful learning to occur in a multimedia environment, the learner must engage in five cognitive processes: (1) selecting relevant words for processing in verbal working memory, (2) selecting relevant images for processing in visual working memory, (3) organizing selected words into a verbal model, (4) organizing selected images into a pictorial model, and (5) integrating the verbal and pictorial representations with each other and with prior knowledge (Mayer, 2002).

In order to design effective multimedia-based instruction, various design principles must be taken into consideration. Principles of multimedia design that have widely been adopted by researchers across subject areas are the seven principles suggested by Mayer (2002). These principles include multimedia principle, spatial contiguity principle, temporal contiguity principle, coherence principle, modality principle, redundancy principle, and individual differences principles. The design principle that this study focused on was the redundancy principle, which is one of the principles that have largely been employed in the design of multimedia-based instruction.

The redundancy principle states that learners can learn better and more deeply just with pictorial information (such as animation) and verbal information (such as narration). The visual text information (i.e., on-screen text), which is presented simultaneously to the verbal information, becomes a redundant material. With this redundant material eliminated, learners will be able to learn better. It is hard for people to focus their attention to all three sources of information at the same time during a presentation (Hoffman, 2006; Mayer, 2002).

Background about Second Language Learners

The term second language is defined by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) as “a language acquired by a person in addition to his mother tongue” (Cook, 2001, p. 13). Second language is typically used as a cover term to refer to any language other than the first language learned by a given learner or group of learners, regardless of the purpose, the type of learning environment (formal or informal), or the number of other non-native languages previously learned or acquired by the learner (Cook, 2001; Smith, 1994). This includes foreign languages, e.g., English as a foreign language for Thais, and languages which
are not the learner’s mother tongue but are widely used in his own community, e.g., English for Spanish-speaking Mexicans. Smith (1994) asserts additionally that the term “second” has a more neutral sense than definitions of terms such as “secondary” or “non-native” which imply lower status. The term “second language” is often abbreviated to “L2” as opposed to “L1” — the mother tongue. However, in certain circumstances, more precise terms, such as L3 and L4, could also be used as in “the influence of a learner’s L2 German upon her L3 Dutch,” for example.

Factors Affecting Second Language Learning

Children begin acquiring their first language (L1) from birth and are surrounded by speakers of the language throughout their development. This contributes significantly to their remarkable success in mastering the language. Second language (L2) learners, on the other hand, can be of any age and in any environment, including places where the language being learned is not a dominant language. While all first language learners are totally successful, it is common for second language learners to fail to fully acquire the language they are learning in addition to their first language (Han, 2004; Robertson, 2008; White, 1989).

Many factors are at work when it comes to second language learning. Some of the most salient factors that are believed to crucially influence second language learning and differentiate second language learners from children acquiring their first language include age of acquisition, mother tongue, fossilization, and learning environment (Lightbown & Spada, 1999; Robertson, 2008; White, 1989).

Age of acquisition. The relationship between age and the potential for success in language acquisition has been documented extensively (Nunan, 1999). It has been widely observed that when young children immigrate to a country where the dominant language is different from their native language, they will manage to successfully learn the new language and are able to eventually speak the language with native-like fluency. An equally high level of proficiency, however, is rarely achieved by their parents. Many adult second language learners have the ability to communicate successfully. Nonetheless, the differences of accent, word choice, or grammatical features often distinguish them from native speakers (Lightbown & Spada, 1999; White, 1989). It is believed that there is a critical period for language acquisition. According to the critical period hypothesis (CPH), there is a time in human development when full native competence is possible when
acquiring a language. The critical period is often claimed to end sometime around puberty. After this period, the innate language acquisition capacities, which are believed to contribute to language acquisition in early childhood, are no longer available. Language learning that takes place after the critical period ends will only rely on general learning abilities, which are the same ones used when learning other kinds of skills or information (Birdsong, 1999; Ipek, 2009; Lightbown & Spada, 1999; Nunan, 1999; Strozer, 1994; White, 1989).

**Mother tongue.** Mother tongue interference is another crucial factor believed to prevent second language learners from mastering the target language. Unlike children learning a first language, all second language learners already have at least one language present in their minds (Cook, 2001), and oftentimes the learners' knowledge of their mother tongue appears to complicate the process of second language acquisition (Bhela, 1999; Ipek, 2009; Swan, 2008). However, it is important to note that this might not be the case for all second language learners. It is often suggested that second language learners whose mother tongue is related to or has elements in common with their second language have an advantage over learners whose native language differs from the language they are learning (Cook, 2001; White, 1989).

**Fossilization.** Many adult second language learners find themselves unable to progress through the learning process even after extended exposure to the target language (Han, 2004). "Fossilization" is the term used to refer to the "phenomenon of non-progression of learning despite continuous exposure to input, adequate motivation to learn, and sufficient opportunity for practice" (Han, 2004, p. 13). Fossilization can be observed even in the second language performance of competent speakers, especially when they are tired or under pressure (Lightbown & Spada, 1999). According to White (1989), fossilization generally involves the use of forms descended from the learner's native language.

**Learning environment.** First language acquisition always takes place naturally with real-life input. When children learn their first language, they are highly motivated to communicate and usually surrounded by native speakers of the language (Robertson, 2008; White, 1989). Moreover, learning, for the most part, occurs informally in learning environments where learners are given opportunities to be exposed to the language without strong pressure to speak fluently and accurately when they are not ready (Lightbown & Spada, 1999). This is strikingly different from learning environments of second language learners where most of the learning takes place. The different between first and second language learners' exposure
to the target language can be observed not only in amount but also in the nature of the language used. Second language learners generally have to follow a syllabus in which language is believed to be learned in an orderly fashion. Their learning is often controlled by the topics and structures of a textbook (Robertson, 2008), and they are often in situations that require "complex language" and "the expression of complicated ideas" (Lightbown & Spada, 1999, pp. 60-61).

Second Language Acquisition Theory

Stephan Krashen (1982) proposed a theory of second language acquisition that has become very influential in the field of second language teaching. Krashen’s theory of second language acquisition consists of five main hypotheses: (a) the acquisition-learning distinction, (b) the natural order hypothesis, (c) the monitor hypothesis, (d) the input hypothesis, and (e) the affective filter hypothesis. Among these is the input hypothesis that is considered to be the central part of an overall theory of second language acquisition (Krashen, 1985a). The input hypothesis model claims that humans acquire language in only one way—by understanding messages or by receiving comprehensible input (Krashen, 1985a, 1985b). Humans naturally focus on the meaning, rather than on form or grammar, of the message as they proceed in their second language development (Cook, 2001; Neuman & Koskinen, 1992).

Comprehensible Input and Second Language Learning

When it comes to second language teaching, the main focus should be placed on the importance of comprehension since acquisition is all about the learner trying to understand. Therefore, one of the most crucial tasks for the language teacher is to provide appropriate and meaningful messages for the student to understand the parts that are beyond their language knowledge (Cook, 2001). The input hypothesis explains why images and other realia are considered extremely valuable to the language teacher, especially in early stages. This is because they provide context helpful for acquiring the target language, and that helps make input comprehensible (Krashen, 1985b).

Neuman and Koskinen (1992) conducted a study of which the results suggested that second language learners developed word meanings and language through comprehensible input. The researchers found that captioned
television as comprehensible input was shown to provide a rich language environment enabling students to acquire new words through context as they developed concepts in science. Precisely, their findings indicated that students who viewed captioned television not only outperformed those who did not on all measures of word knowledge, but they also appeared to remember more science content than their counterparts. This demonstrated, therefore, that communicating messages to second language learners through different modalities appeared to enhance their learning rather than negatively affecting their attention capacity (Neuman & Koskinen, 1992).

PURPOSE OF THE STUDY

The purpose of this study was to investigate the effects of different multimedia learning environments on the learning outcomes of second language learners. Specifically, this study assessed Mayer’s redundancy principle by comparing the learning outcomes of two groups of students presented with two formats of a multimedia lesson: (1) images with audio and (2) images with audio and on-screen text.

RESEARCH QUESTION

This study sought to answer the research question: Do second language learners who are presented with a multimedia lesson containing only images and audio differ in subsequent comprehension performance from second language learners who are presented with a multimedia lesson containing images, audio, and on-screen text?

PARTICIPANTS

The sample for this study consisted of 230 undergraduate students enrolled in a single English course at a university in Thailand. The participants were recruited through nonprobability convenience sampling and were a sample of the entire undergraduate student population at this particular university of approximately 25,000. Although the participants were selected based on their convenience and availability, they were required to meet certain criteria of the study. First, all participants were in the same age range, which was between 18 and 21 years. Second, they were enrolled in at least
one English course at the time the study was conducted. Finally, to minimize effects due to potential differences among academic disciplines, all participants were in the same academic discipline. The researcher contacted the director of the university language institute, who was in charge of all English courses offered at the university, to ask for permission to conduct the study. Once the permission was granted, the researcher contacted English language instructors whose students met all of the criteria of the study. Permission was requested from each instructor to access her or his course section and request participation from the students.

Nine sections of students, with approximately 25-35 students each, were recruited. Students within each section remained together as intact groups, but these nine sections of students were randomly assigned into two different groups defined by the type of multimedia learning environment that the students experienced. There were 113 students participating in Group 1 (images with audio only) and 117 students participating in Group 2 (images with audio and on-screen text). Relevant demographic information of the participants, which includes gender, age, and English proficiency, was also collected (Tables 1 and 2).

As shown in Table 1, the sample consisted of 230 participants, including 159 females (69.87%) and 71 males (30.13%). The Images with Audio Group consisted of 79 females (69.91%) and 34 males (30.09%) while the Images with Audio and On-Screen Text Group consisted of 80 females (68%) and 37 males (31.62%).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Images + Audio</th>
<th>Images + Audio + On-Screen Text</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>79</td>
<td>69.91%</td>
<td>80</td>
</tr>
<tr>
<td>Male</td>
<td>34</td>
<td>30.09%</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td>100.00%</td>
<td>117</td>
</tr>
</tbody>
</table>

Table 2 provides descriptive statistics for the overall age and English proficiency of the participants. The mean age of all participants was 19.13 years ($SD = 0.50$). The mean age of the participants in the Images with Audio Group was 19.16 years ($SD = 0.49$) and the mean age of the participants in the Images with Audio and On-Screen Text Group was 19.10 years ($SD = 0.50$).
The average English proficiency score of all participants was 55.37 ($SD = 13.47$). The mean English proficiency score of the participants in the Images with Audio Group was 54.75 ($SD = 13.66$) and the mean English proficiency score of the participants in the Images with Audio and On-Screen Text Group was 55.97 ($SD = 13.31$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Images + Audio</th>
<th>Images + Audio + On-Screen Text</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>Mean</td>
<td>$SD$</td>
</tr>
<tr>
<td>Age</td>
<td>113</td>
<td>19.16</td>
<td>0.49</td>
</tr>
<tr>
<td>English Proficiency</td>
<td>113</td>
<td>54.75</td>
<td>13.66</td>
</tr>
</tbody>
</table>

Review of the data shows that the English proficiency scores of participants in the Images with Audio Group is close to normally distributed.

**INSTRUMENTATION AND MATERIALS**

Two sets of instrumentation, which consisted of learning materials and tests, were used in this study.

**Learning Materials**

To achieve the goal of this study, two different multimedia learning environments were developed: (1) images with audio and (2) images with audio and on-screen text. The content presented in both multimedia learning environments was exactly the same. In this study, the two multimedia learning environments were two different formats of a narrated video, one with on-screen text and one without on-screen text. The video was in English, and the content of the video consisted of two related science topics: natural selection and evidence for evolution. Both formats of the video were 20 minutes in length. This video was, in fact, a combination of two videos originally created by Stated Clearly, a group of artists, scientists, and educators who have come together to create a series of short animations to teach science in a simple, friendly manner. Validity evidence for the content presented in the videos has been obtained using experts in the field. The permis-
sion to use their videos as part of this study has been granted by Jon Perry, the founder of Stated Clearly. The combined version of the two videos both with and without captions was installed into computers in a computer lab where the study took place. No internet access was required for participants to access the learning materials.

Pretest and Posttest

A pretest with 20 test items was developed by the researcher to be used as a tool to assess the participants' prior knowledge of the learning content to be presented to them. The same set of 20 test items also was used as a posttest to measure the participants' learning achievement. This was to ensure that the scales used on both the pretest and the posttest were similar. However, the sequence of the test items in the pretest was different from that of the posttest to minimize potential order effects. Both the pretest and the posttest were paper based, and they were mixed-format tests, consisting of different types of test items, which included multiple choice, short answer, fill-in-the-blank, true-false, and matching.

Content and Instrument Validation

To ensure content accuracy and appropriateness of the tests, a group of experts including two second language instructors, two second language learners, and a scientist were invited to review the test items. The participants were also asked to take the pretest before watching the video and to take the posttest after they finished watching the video. A Wilcoxon signed-ranks non-parametric test (used because of the small sample size and lack of normality of the difference scores) showed that participants scored significantly higher on the posttest than the pretest ($z = 2.023$, $p = .043$). The effect size for the difference was large ($r = .90$). This provides some supporting evidence for the validity of scores obtained from the test. However, some of the test items required revision because they appeared to be too complicated, as none of the participants was able to respond correctly even though the specific content knowledge needed was clearly presented in the video. Some of the test items, on the other hand, appeared to measure extraneous knowledge, as participants could simply use prior, common knowledge to respond to those items correctly without having to watch the video. The tests used in this study were revised based on this process of content and instrument validation.
DATA COLLECTION PROCEDURES

A list of participants of both groups—(1) images with audio and (2) images with audio and on-screen text—was prepared prior to data collection. The intervention took place in a computer lab where the researcher was available at all times. Students within the same sections assigned by the university participated in the study at the same time, and all of them received the same multimedia instructional materials presented in two different formats, which were a video containing images with audio and a video containing images with audio and on-screen text. Altogether, there were nine sections of students, randomly assigned into two groups. Approximately half of the students (N = 113) participated in a multimedia lesson where the learning content was presented through images and audio, and the other half of the students (N = 117) participated in a multimedia lesson where the same learning content was presented through images, audio, and on-screen text.

The researcher began by informing the participants of the purpose of the study and the steps they were about to go through. Then, each of the participants received a small card with his/her name and English proficiency test score on it. Participants’ names were present on the cards only to ensure that the correct English proficiency test score was given to the right participant. Participants’ names, however, were not used or present anywhere else during or after the intervention. A space for participants to fill in their participant code was also provided on the card. After giving each participant a card with his/her information on it, the researcher randomly distributed a paper-based pretest to all of the participants. Each copy of the pretest was labeled with a participant code. Then the researcher instructed participants to copy the participant code from the test to their card so that they could put this same code on the posttest, which was given to them afterward. Participants were also asked to fill in their demographic information, which included their gender, age, and English proficiency test score. Once the participants completed the pretest, the researcher collected the completed test from the participants. Then the participants began their multimedia lesson previously installed on the computers. Participants were not allowed to take notes. As soon as the lesson was complete, a paper-based posttest was administered to assess the participants’ learning outcomes. Participants were asked to write their participant code in the box located on the top right corner of the test. This allowed the researcher to anonymously match participants’ pretests to their posttests.
DATA ANALYSIS

The independent variable of interest in this study was the intervention, with two levels of multimedia learning environment. English proficiency scores and gender were used as control variables. The dependent variable for this study was learning outcomes as indicated by growth scores from pretest to posttest. Means, standard deviations, and ranges for pretest scores, posttest scores, growth scores, and English proficiency scores were computed using SPSS. The reliability of the pretest and posttest scores was also estimated by computing KR-20. A repeated-measures analysis of covariance (ANCOVA) was carried out to (1) evaluate change between pretest and posttest, (2) evaluate mean differences between the two groups, and (3) evaluate group differences in growth from pretest to posttest, using gender and English proficiency as control variables. Additionally, an analysis of covariance was performed to assess group differences in posttest scores, controlling for pretest scores, gender, and English proficiency.

A power analysis indicated that, assuming a moderate effect size for the difference between groups in the population, and using alpha = .05, a total sample size of $N = 128$ participants would be required for 80% power. Power analysis also indicated that, assuming a large effect size for the difference between groups in the population, and using alpha = .05, a total sample size of $N = 52$ participants would be required for 80% power. Table 1 provides the details of these analyses. Eta-squared ($\eta^2$) was used to determine the effect size. After the data analysis was complete, the results of this study were discussed, and an interpretation of the data was written.

STATISTICAL ANALYSIS

Prior to analyses, internal consistency estimates of reliability (Cronbach's coefficient alpha) were calculated for the pretest and the posttest used in this study and descriptive statistics for the entire sample were computed. Cronbach's alpha was .51 for the pretest and .64 for the posttest. Table 3 reports the means, standard deviations, and skewness of pretest, posttest, and growth scores of participants in the Images with Audio Group and the Images with Audio and On-Screen Text Group. The overall mean, standard deviation, and skewness of pretest, posttest, and growth scores also are reported.
Further review shows that the pretest scores of participants in the Images with Audio Group is close to normally distributed. The pretest scores of participants in the Images with Audio and On-Screen Text Group, however, appears to be negatively skewed. Histogram for pretest scores of participants in the Images with Audio and On-Screen Text Group.

**RESEARCH QUESTION AND RESULTS**

Do second language learners who are presented with a multimedia lesson containing only images and audio differ in subsequent comprehension performance from second language learners who are presented with a multimedia lesson containing images, audio, and on-screen text?

A repeated-measures analysis of covariance (ANCOVA) was conducted to evaluate change in comprehension performance from pretest to posttest, mean differences in comprehension performance between the two groups, and group differences in comprehension growth from pretest to posttest. Additionally, gender was employed as a factor and English proficiency score as a covariate. In addition, analysis of covariance was also carried out to assess group differences in posttest scores, controlling for pretest scores, gender, and English proficiency. Eta-squared ($\eta^2$) and Cohen’s $d$ were used to compute the effect size.

For the repeated-measures ANCOVA analyses, which used the pretest and posttest comprehension scores as the paired outcomes, Box’s test of equality of covariance matrices indicated that covariances were equal across groups, $p = .392$. Table 4 shows the results for the ANCOVA. The results indicated that, across groups, a statistically significant change between pretest to posttest occurred, $F(1, 225) = 3.98, p = .047$. Computation of eta-squared suggested a small effect size ($\eta^2 = 0.01$); however, this statistic was likely influenced by sample size and reflected the large amount of unexplained error variation in the data. Cohen’s $d$, which reflects a standardized index of mean growth as opposed to variance accounted for, was
$d = 2.09$, indicating a very large increase in comprehension from pretest to posttest. The observed growth from pretest to posttest, however, did not significantly differ between the two treatment groups, $F(1, 225) = 0.23$, $p = .635$ (Table 5). Examination of the data shows that participants in both the Images with Audio Group and the Images with Audio and On-Screen Text Group appear to display a similar amount of growth. Also, there was no significant difference in growth between males and females, $F(1, 225) = 0.58$, $p = .449$. However, the results suggest that there was a significant three-way group $\times$ gender $\times$ time interaction effect, $F(1, 225) = 5.79$, $p = .017$. The effect size was $\eta^2 = 0.02$, however, suggesting a small effect. Analysis shows that males appear to show greater comprehension growth in the Images with Audio Group than in the Images with Audio and On-Screen Text Group. In contrast females tend to show greater growth in the Images with Audio and On-Screen Text Group.

### Table 4
Results for the Repeated-Measures Analysis of Covariance (ANCOVA) of Comprehension Across Time by Group and Gender

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>16.818</td>
<td>1</td>
<td>16.818</td>
<td>3.984</td>
<td>.047</td>
</tr>
<tr>
<td>Time $\times$ English Proficiency</td>
<td>163.514</td>
<td>1</td>
<td>163.514</td>
<td>38.731</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Time $\times$ Group</td>
<td>.952</td>
<td>1</td>
<td>.952</td>
<td>.225</td>
<td>.635</td>
</tr>
<tr>
<td>Time $\times$ Gender</td>
<td>2.426</td>
<td>1</td>
<td>2.426</td>
<td>.575</td>
<td>.449</td>
</tr>
<tr>
<td>Time $\times$ Group $\times$ Gender</td>
<td>24.441</td>
<td>1</td>
<td>24.441</td>
<td>5.789</td>
<td>.017</td>
</tr>
<tr>
<td>Error(Time)</td>
<td>949.906</td>
<td>225</td>
<td>4.222</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5
Test of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1140.817</td>
<td>1</td>
<td>1140.817</td>
<td>139.683</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>English Proficiency</td>
<td>401.675</td>
<td>1</td>
<td>401.675</td>
<td>49.181</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Group</td>
<td>8.088</td>
<td>1</td>
<td>8.088</td>
<td>.990</td>
<td>.321</td>
</tr>
<tr>
<td>Gender</td>
<td>.419</td>
<td>1</td>
<td>.419</td>
<td>.051</td>
<td>.821</td>
</tr>
<tr>
<td>Group $\times$ Gender</td>
<td>5.649</td>
<td>1</td>
<td>5.649</td>
<td>.692</td>
<td>.406</td>
</tr>
<tr>
<td>Error</td>
<td>1837.623</td>
<td>225</td>
<td>8.167</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The second analysis employed ANCOVA, with the posttest comprehension scores as the outcome. Levene’s test was not statistically significant, \( p = .380 \), indicating equal variances across groups. As displayed in Table 6, a test of between-subject effects indicated that group and gender did not have statistically significant effects on the posttest scores, \( F(1,224) = 0.03, p = .854 \), and \( F(1,224) = 0.23, p = .631 \), respectively. However, there was a significant, disordinal group \( \times \) gender interaction effect on the posttest scores, \( F(1,224) = 5.81, p = .017 \). The effect size was \( \eta^2 = 0.02 \), suggesting a small effect. Analysis of the data shows that males tend to perform better on the posttest under the Images with Audio condition while females tend to perform better under the Images with Audio and On-Screen Text condition.

Table 6

<table>
<thead>
<tr>
<th>Source</th>
<th>Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>710.430</td>
<td>5</td>
<td>142.086</td>
<td>24.866</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intercept</td>
<td>300.833</td>
<td>1</td>
<td>300.833</td>
<td>52.648</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>English Proficiency</td>
<td>455.320</td>
<td>1</td>
<td>455.320</td>
<td>79.685</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prettest Total</td>
<td>144.536</td>
<td>1</td>
<td>144.536</td>
<td>25.295</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Group</td>
<td>0.194</td>
<td>1</td>
<td>0.194</td>
<td>0.034</td>
<td>.854</td>
</tr>
<tr>
<td>Gender</td>
<td>1.323</td>
<td>1</td>
<td>1.323</td>
<td>0.232</td>
<td>.631</td>
</tr>
<tr>
<td>Group ( \times ) Gender</td>
<td>33.184</td>
<td>1</td>
<td>33.184</td>
<td>5.807</td>
<td>.017</td>
</tr>
<tr>
<td>Error</td>
<td>1279.935</td>
<td>224</td>
<td>5.714</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>46068.000</td>
<td>230</td>
<td>5.714</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1990.365</td>
<td>229</td>
<td>5.714</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \text{\textit{a. R-Squared = .357 (Adjusted R-Squared = .343)}} \)

**DISCUSSION OF FINDINGS**

Participants’ prior knowledge of the learning content was measured by a pretest examination before they engaged with one of the two multimedia learning environments. Then, a posttest examination was administered to assess their learning achievement. To answer the research question of whether second language learners who participated in a multimedia lesson containing only images and audio differed in subsequent comprehension performance from those who participated in a multimedia lesson containing images, audio, and on-screen text, their growth scores, which were the differences between pretest and posttest scores, were compared.
According to the redundancy principle, one of the design principles under the cognitive theory of multimedia learning (Mayer, 2002), a better way for people to learn is from pictorial information, or images, and verbal information, or audio. Visual text information, or on-screen text, which is presented simultaneously to the verbal information, becomes a redundant material and should be eliminated. Mayer (2002) explains that cognitive overload can occur in the visual/pictorial channel when both images and text are presented visually at the same time. On the other hand, when only pictorial information and verbal information are presented, the pictorial information enters through the eyes and is processed in the visual/pictorial channel while the verbal information enters through the ears and is processed in the auditory/verbal channel. The learner is, therefore, believed to be able to engage in appropriate cognitive processing and learn more effectively (Clark & Mayer, 2008; Mayer, 2002). Previous research on multimedia learning has also shown to support this claim and proved Mayer’s redundancy principle to be accurate. Acha (2009) and Moussa-Inaty and Atallah (2012) found in their studies that working memory appeared to be overloaded due to redundancy when images and written text were presented simultaneously on screen, and this appeared to have a negative impact on learning performance.

The results of this study, however, showed that a statistically significant change from pretest to posttest occurred for the combined Images and Audio and Images with Audio and On-Screen Text Groups. Participants in both groups also appeared to display a similar amount of growth. Although not statistically significant, at the sample level, participants in the Images with Audio and On-Screen Text were found to have a slightly higher amount of growth than those in the other group, \( p = .635 \). These non-significant results appeared to contradict the redundancy principle because participants in the group where a multimedia lesson was presented through images, audio, and on-screen text were able to perform as well as those in the group where the same multimedia lesson was presented without the on-screen text element. These findings, however, are in line with those obtained from a number of previous studies in the field of multimedia learning (Akbiyik & Akbiyik, 2010; Karakas & Saricoban, 2012; Moussa-Inaty & Atallah, 2012; Winke, Gass, & Sydorenko, 2010; Yuksel & Tanriverdi, 2009). Participants’ learning outcomes were found to significantly improve whether or not visual text was provided while participating in a multimedia lesson (Karakas & Saricoban, 2012; Yuksel & Tanriverdi, 2009). Their results, as well, did not indicate a significant difference between the two groups; however, this suggests that providing visual text on screen in addition to images and audio did not appear to negatively impact the learners.
Winke, Gass, and Sydorenko (2010) investigated the effects of captioning during video-based listening activities, and they found that having both images and written text together on screen appeared to enhance the learning outcomes of foreign language learners. The written text, which was presented in the form of captions in their videos, appeared to be a crucial element and was shown to be beneficial to students as it enabled them to understand the videos better. This is consistent with what was found in a study conducted by Akbiyik and Akbiyik (2010). In their study, the researchers investigated whether different types of multimedia presentation affected students' interpretation achievement, and it was found that students who learned through multiple modes of learning were able to perform significantly better than those who learned through only a single input mode. Results showed that students with highest achievement mean scores were those in the group where text and audio were presented together. Similarly, Moussa-Inaty and Atallah (2012) also found that students who were given materials that included the reading component, or written text, appeared to outperform students who were given materials that contained only the listening and/or graphics components.

The fact that providing all three types of information—pictorial information, verbal information, and visual text information—did not appear to be overwhelming to the learners in this study calls into question the validity of Mayer’s redundancy principle with second language learners because, if Mayer were right, the learners who were presented with a multimedia lesson containing images, audio, and on-screen text would have performed worse than those who were presented with a multimedia lesson containing only images and audio. Results from this study, on the other hand, are consistent with the theory of second language acquisition, particularly the input hypothesis (Krashen, 1982). According to the input hypothesis, receiving comprehensible input is crucial for humans to acquire language (Ipek, 2009; Johnson, 2004; Krashen, 1985a). That is, learners need to be able to understand the message being conveyed to them; therefore, it is tremendously helpful for second language learners to have written materials to supplement the aural media as this helps make the aural input comprehensible (Krashen, 1985b). Examples of comprehensible input could include television with subtitles or videos with captions containing messages understandable to the learners (Neuman & Koskinen, 1992). Since language learners often find audiovisual input in a foreign language challenging and difficult to comprehend (Danon, 2004), the multimedia lesson containing the on-screen text element used in the present study can therefore be considered as comprehensible input because, with on-screen text provided, learners are able to vi-
ualize what they hear (Danan, 2004). This can also help learners to be more certain of ambiguous input and to analyze words and phrases, which can lead to a better understanding of the content being presented (Chai & Erlam, 2008; Danan, 2004; Winke, Gass, & Sydorenko, 2013).

It is important to note, however, that although cognitive overload was not observed when on-screen text was provided additionally to images and audio the multimedia learning environment in this study, more evidence would be needed in order to draw a firm conclusion that cognitive overload did not occur at all during the learning process. It is possible that there might have been some degree of cognitive overload taking place, which could have contributed to the lack of significance in this study. Results from this study also reveal a significant gender difference. Interestingly, females were found to perform significantly better than males in the multimedia learning environment where on-screen text was present, whereas males, on the other hand, were found to perform significantly better than females in the multimedia learning environment where on-screen text was not available. In other words, female students in this study appeared to benefit more from the presence of on-screen text than males. Cognitive overload could have played a role here too, but more research would be needed to prove this. Another possible explanation for this gender difference could be that males and females are different in terms of information processing (Darley & Smith, 1995; Meyers-Levy & Sternthal, 1991). Males tend to be selective and do not often engage in comprehensive processing of all information available to them. In contrast, females tend to be comprehensive and engage in comprehensive analysis of all available information (Meyers-Levy & Sternthal, 1991).

**IMPLICATIONS**

The role of multimedia technology in education is unquestionably remarkable because with multimedia integrated, a dull lesson can be transformed into a more lively and engaging one (Chuang & Liu, 2012; Pun, 2013). Designing an effective multimedia learning environment can, however, be challenging. Based on his extensive research on multimedia learning, Mayer (2002) has suggested a number of helpful multimedia design principles as part of the cognitive theory of multimedia learning to serve as guidelines for educators who wish to incorporate multimedia into their teaching. Among them is the redundancy principle, a well-documented multimedia design principle. The primary rule according to the redundancy principle is to never include on-screen text in addition to images and audio in a multi-
media learning environment because it is not only redundant to the audio but also overwhelming to the visual channel as the learner will need to process both images and on-screen text at the same time (Clark & Mayer, 2008; Mayer, 2002). While the redundancy principle has been supported by a large body of research (Austin, 2009; Kalyuga, Chandler, & Sweller, 1999, 2000; Mayer, Heiser, & Lonn, 2001; Mousavi, Low, & Sweller, 1995), findings from the present study suggest otherwise.

As suggested by the results of this study, the presence of on-screen text, although redundant to the information contained in the audio, did not appear to impede learning of second language learners. These results indicate that the redundancy principle may not always hold true. The same multimedia effects, which were originally discovered in the context of science instruction, may not occur when it comes to second language instruction where lessons are delivered in a language foreign to the learners. Factors such as learning context and learners' needs should be taken into consideration when designing a multimedia learning environment because a multimedia learning environment that can successfully enhance learning at one place may not do quite well at another due to different learning circumstances. It is, therefore, important that multimedia learning environments are custom made to fit the target groups of learners so that they can better accommodate the learners' needs and thereby effectively facilitate learning. Results obtained from this study suggest that including on-screen text in a multimedia learning environment developed for second language learners is beneficial to them because when the learning content is in a foreign language and is presented to them solely through audio, learners may encounter difficulties in comprehending the information. Having on-screen text available to the learners is, therefore, advantageous as it can help them make sense of what they hear, which can lead to greater comprehension and more favorable learning outcomes.

References


Examining the Effects of Different Multimedia Learning Environments


