December 2019

Multimedia Input Modes, the Modality Principle, and the Redundancy Principle for University ESL Students' Learning

Yinan Liu
Syracuse University

Abstract

This study compared three multimedia input modes in the modality and redundancy principles (Mayer, 2009) in terms of university ESL (English as a Second Language) students' learning and examined the applicability of the modality and redundancy principles for ESL students. Mayer's modality and redundancy principles (2009) inform the design of effective multimedia lessons. However, the two principles originally stemmed from experimental studies examining students' learning in their native language and did not include ESL students in the discussion. Based on the modality and redundancy principles, added on-screen text and graphics lead to an overload in learner's visual channel, which undermines learning (Clark & Mayer, 2011).

For ESL students' multimedia learning, the cognitive theory of multimedia learning (Mayer, 2009) suggests that on-screen text in the input modes of graphics + text and graphics + audio + text might overload the visual channel to impede learning. However, according to the cognitive load theory (Sweller, 2014), text might also reduce the processing demands for identifying and decoding auditory input to facilitate learning. Due to the limited number of empirical studies, it was inconclusive if verbatim text aids or hinders ESL students' learning, and it was unclear if the modality and redundancy principles apply for ESL students.

An initial Study addressed common validity issues, such as lack of control of instruments and materials, in related studies, and it quantitatively tested the applicability of the modality and redundancy principles for ESL students' learning. Both knowledge retention and vocabulary test results indicated that input modes did not have an impact on ESL students' learning, and consequently the modality and redundancy principles were insignificant. An additional study,

Study 2, addressed the implementation issues and limitations of Study 1 to provide more rigorous findings.

Based on the findings of both Study 1 and Study 2, the modality and redundancy principles did not apply for ESL students' content knowledge and vocabulary learning when certain multimedia learning principles were followed. Both Study 1 and Study 2 extended Mayer's modality and redundancy principles by examining their applications to ESL students, as well as provided empirical evidence for designing effective multimedia instruction for ESL students.

Multimedia Input Modes, the Modality Principle, and the Redundancy Principle for University ESL Students' Learning

By

Yinan Liu

B.A., Dalian University of Foreign Languages, 2003

M.A., Syracuse University, 2009

Dissertation

Submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Literacy Education

Syracuse University

December 2019

Acknowledgements

I am grateful for the help and guidance I received from my committee members: Dr. Zaline Roy-Campbell, Dr. Rachel Brown, and Dr. James Bellini. This dissertation would not be possible without their help.

I am most grateful for the love and supports of my family: my wife, Xin, my three beautiful daughters, Changchang, Mimi, and Lily. You are all the beauties in the world.

I dedicate this study to all the mavericks in the field of English education.

Table of Contents

List of Illustrative Materials

TABLES

FIGURES

CHAPTER 1

INTRODUCTION

The chapter begins with the background to the study and the problem statement. It then identifies the research questions, followed by definitions of key terms.

Background

For years, I, as an English learner, have been learning English knowledge, such as new vocabulary, from watching my favorite captioned movies and TV shows. I find that my interest in the movie encourages me to figure out the language that I do not know. Additionally, when encountering an unfamiliar word in the movie, I find that graphics provides the social and cultural contexts and teaches the meaning of the word, narration teaches the pronunciation of the word, and on-screen text or caption teaches the spelling of the word. For me, a captioned movie is a multimedia vocabulary lesson built on my interest, and it is able to address all four aspects of vocabulary learning, which are shape, meaning, sound, and use (Nation, 2011).

For years, I, as an English and Chinese language instructor, have integrated captioned movies or TV shows into my syllabi to teach language skills. On the end of year evaluation form, a great majority of my students reported that they preferred movies to textbooks to learn the language, because they were able to learn authentic language and culture in a relaxing and engaging learning environment. Additionally, most students commented that caption was very important in helping them understand the video.

In an English Communications class, I showed a five-minute captioned video clip to the class. The clip was about the formation of El Nino, which was different from the usual movie clip I let the class watch. The clip was more educational. It consisted of graphics depicting how

El Nino influenced the environment and alternating graphics of a teacher explaining the mechanism of El Nino (see Figure 1). After watching the clip, to my surprise, no one could recall what El Nino was. Several students were able to recognize a few terms and mostly from their prior knowledge. What was more intriguing and unexpected was that almost half of the class commented that the added caption was distracting because they did not know where to look, which contradicted their previous unanimous support for captioning. Some students also reported that the teacher in the video did not do a good job explaining the phenomenon, and they could have learned better if there had been an animation explaining El Nino.

Figure 1. A screenshot of the clip about El Nino.

Students' responses to the clip led me to questions:

1. Why did captioning seem to be favored in some videos but not in others?
2. Did graphics of a teacher promote learning when he or she was teaching content knowledge in a video?
3. What factors need to be taken into account when teachers choose or design instructionally appropriate videos for ESL students?

These questions sparked my interest in exploring captioning or on-screen text for ESL students' learning, as well as other strategies to promote ESL students' learning via multimedia.

After an initial review of the literature on this topic, I learned that:

1. Videos consisting of graphics and words can be seen as multimedia, because information in a video is conveyed via two or three media (Mayer, 2009).
2. Caption is the verbatim transcription of narration when narration is present. Captioning was reported to be generally helpful as suggested by Montero Perez, Van Den Noortgate and Desmet (2013) in their meta-analysis study about captioning for ESL students, which could be translated that students preferred graphics + audio + text to graphics + audio.

Searching studies regarding multimedia and on-screen text, I found that Mayer's modality and redundancy principles (2009) examined the role of on-screen text in multimedia learning and Mayer's other research-based learning principles directed the design of effective multimedia lessons. Based on the redundancy principle (Mayer, 2009), students learn better from graphics + audio than from graphics + audio + text, which contradicts my students' responses to on-screen text. I suspected that the redundancy principle might not apply for ESL students' learning, which led to this study to identify empirical evidence to support my suspicion.

Problem Statement

Currently, multimedia technologies and tools, such as computer games and PowerPoint slides, are ubiquitously utilized in university classes to enhance content knowledge learning (Borgh & Dickson, 1992; Doty, Popplewell, & Byers, 2001; Goldenberg, Heinze, & Ba, 2004; Rose & Meyer, 2002). Studies reported that university students found multimedia lessons to be both interesting and effective (e.g. Shuell & Farber, 2001). Technological advances enable instruction featuring multimedia to be more accessible and popular in university classrooms (Goldenberg, Heinze, & Ba, 2004).

To explain and direct the design of effective multimedia lessons, Richard Mayer (2009) proposed the cognitive theory of multimedia learning and a set of multimedia learning principles. Multimedia lessons feature different combinations of words and graphics (Mayer, 2009, 2014). Such combinations are input modes through which new information is conveyed and taught. The modality principle and the redundancy principle (Mayer, 2009) are two fundamental principles that focus on input modes. According to the modality principle, the input mode of graphics + audio is more effective for learning than that of graphics + text. According to the redundancy principle, the input mode of graphics + audio is more effective for learning than that of graphics + audio + text. Text and narration that teach the content contain the same verbal information. Unlike on-screen keywords, text consists of complete sentences. Both principles suggest that text presented with graphics could lead to a visual channel overload, which hampers learning (Mayer, 2009).

Although a substantial body of studies are concerned with multimedia learning principles, most of these studies examined teaching new content when verbal information is in the learner's native language (Mayer, 2009, 2014a). Empirical studies on multimedia lessons in learner's second language are scarce (Mayer, 2009, 2014a; Mayer, Lee, & Peebles, 2014; Sweller et al., 2011). The modality and redundancy principles discussed native English-speaking students' learning but failed to address the increasing number of English-as-a-second-language (ESL) students in US university classrooms. For ESL students, on-screen text might promote learning by reducing the processing load from identifying and decoding unfamiliar English narration (Meskill, 1996; Winke, Gass & Sydorenko, 2010), which contradicts the detrimental role of text suggested in the modality and redundancy principles. Therefore, the modality and redundancy principles might not apply to ESL students' learning.

What defines ESL students is their lack of English proficiencies, and most empirical studies on ESL students' learning via multimedia investigate their development of language skills. For example, a plethora of studies (e.g. Aldera & Mohsen, 2013; Bird & Williams, 2002; Markham, 1999) addressed the influences of redundant verbatim on-screen text on ESL students' language development. However, ESL students come to US universities to learn more than language skills. More importantly, they learn content knowledge in specific disciplines with their native English-speaking peers in the same class. ESL students in US classrooms lack the linguistic and cultural knowledge of native speakers (Shatz & Wilkinson, 2010). They are likely to be disadvantaged when presented with multimedia lessons designed for their native English-speaking classmates (Casado & Dereshiwsky, 2001; Huang, 2005, 2006). To ensure that a multimedia lesson is accessible for both native English-speaking and ESL students, multimedia learning principles (Mayer, 2009) that provide useful insights for the design of effective multimedia lessons should also address ESL students' learning.

For ESL students, verbatim on-screen text in a multimedia lesson could be helpful because it reduces the cognitive processing needed for identifying and decoding fleeting auditory input (Clark & Mayer, 2011; Lee & Mayer, 2018; Meskill, 1996). Therefore, the modality and redundancy principles that suggest the detrimental role of verbatim text might not apply for ESL students' learning. There was little empirical evidence to substantiate such an assumption.

Few studies addressed ESL students' learning of content knowledge via multimedia. Even fewer studies focused on different multimedia input modes for ESL students' learning of content knowledge (Mayer, 2009, 2014). The very few studies (Huang et al., 2016; Lin et al., 2016; Mayer, Lee, & Peebles, 2014; Shadiev et al., 2017) on ESL students' content knowledge learning via input modes also suffered several common validity threats, such as lack of language control

of test instruments. Mayer et al. (2014) conducted a study and compared university-level ESL students' content knowledge learning via the input mode of graphics + audio + text and that of graphics + audio. They found no statistical significance in the learning outcomes. However, Mayer et al. (2014) neither examined all three multimedia input modes in the modality and redundancy principles (i.e. graphics + text, graphics + audio, and graphics + audio + text), nor tested the applicability of the two principles for ESL students, which inspired this study.

Research Questions

This dissertation sought to extend the theoretical basis of Mayer's modality and redundancy principles by investigating all three input modes to include ESL students. More specifically, this study examined ESL students' acquisition of new content knowledge in geography from a multimedia lesson about lightning in one of the three input modes. The three research questions are:

1. Which of the three input modes is the most facilitative for university-level ESL students' learning?

   (H_0: ESL students learn the same from graphics + audio, as from graphics + text, and from graphics + audio + text.)
2. Does the modality principle apply to university-level ESL students' multimedia learning?

   (H_0: ESL students learn the same from graphics + audio as from graphics + text.)
3. Does the redundancy principle apply to university-level ESL students' multimedia learning?

   (H_0: ESL students learn the same from graphics + audio as from graphics + audio, + text.)

When I conducted Study 1 (Liu, Jang, & Roy-Campbell, 2018), the findings supported the ones of Mayer et al. (2014) that there was no statistically significant difference in ESL

students' learning outcomes between the graphics + audio group and the graphics + audio + text group. However, I was unconfident of the results due to some implementational and methodological issues that threatened the internal and external validity of the study. I therefore conducted a follow-up study (Study 2) to replicate Study 1 and extend it by addressing its limitations for more rigorous findings. This dissertation will elaborate on the procedure and findings of the two studies and compare their results to provide a deeper understanding of the applicability Mayer's modality and redundancy principles to ESL students.

Chapter 2 provides the theoretical framework and a review of the literature. Chapter 3 describes the methodology of the two studies, with an explanation of the how the second study addressed the limitations of Study 1. Chapter 4 provides the results of Study 2. Chapter 5 compares and discusses the findings of the two studies in relation to extending Mayer's (2009) modality and redundancy principles.

Definitions

Caption

In this study, caption is the synchronous verbatim on-screen text of auditory verbal input usually at the bottom of the screen. As a textual form of verbal input, caption traditionally supports the comprehension of the hearing impaired. Currently, caption is increasingly used for the learning and instruction of ESL students (Bowe & Kaufman, 2001; Evmenova, 2008; Linebarger, 2001).

Text and caption are interchangeable in this study. Text and narration convey the same instructional messages to teach content knowledge. Different from text that consists of complete sentences, on-screen keywords summarize text and are placed adjacent to their corresponding graphics.

ESL Students

In this study, ESL students refer to those who use English as a second language or a non-native language in native English-speaking higher educational institutions. These ESL students do not have comparable English proficiency as their native English-speaking counterparts, and often need support to meet the language demands for academic success (Cheng, Myles, & Curtis, 2004). ESL is sometimes interchangeable with a more accurate term of ESOL or English to Speakers of Other Languages.

Learning

In this study, learning is defined as "a change in the learner's knowledge due to experience" (Clark & Mayer, 2011, p. 32). According to this definition, learning involves addition of new knowledge to learner's knowledge base due to experience, such as a lesson or a treatment. Therefore, this study examines ESL students' learning or acquisition of new content knowledge in geography from a multimedia lesson in one of three input modes.

Learning is different from comprehension in that comprehension does not entail acquisition of new knowledge.

Multimedia Lessons

A multimedia lesson is a lesson taught via multimedia. According to Mayer (2014a), "multimedia instruction consists of instructional messages that contain words (such as printed or spoken text) and pictures (such as illustrations, diagrams, photos, animation, or video)" (p. 385). As a defining characteristic of a multimedia lesson, graphics can be of any imagery forms such as pictures, animations, and illustrations etc.

A multimedia lesson is different from multimedia for other purposes that also combine graphics and words, such as movies and TV shows, in that it requires careful planning and

designing to achieve specific instructional goals.

Multimedia Input Modes

Input modes of multimedia lessons consist of different combinations of graphics, text, and audio through which new messages are represented and shared. Since the presence of graphics is a key characteristic of multimedia instruction (Mayer, 2009), multimedia input modes for instructional purposes in this study are comprised of graphics + audio, graphics + text, and graphics + audio + text.

CHAPTER 2

LITERATURE REVIEW

This chapter provides the theoretical framework, the relevant empirical studies, and addresses the gaps in the literature that led to the research questions.

Theoretical Framework

Since this study attempted to examine the applicability of Mayer's (2009) modality and redundancy principles for ESL students, the theoretical framework that informs this study includes Mayer's (2003) cognitive theory of multimedia learning on which the original modality and redundancy principles were based, the modality and redundancy principles (Mayer, 2009), and the cognitive load theory (Sweller, 2014) that explained ESL students' multimedia learning.

The Cognitive Theory of Multimedia Learning

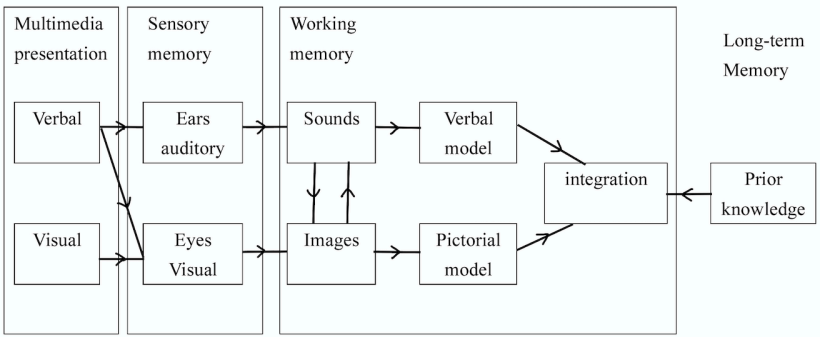

Figure 2. The cognitive theory of multimedia learning (adapted from Mayer, 1997, 2001, 2003, 2005a, 2009).

The cognitive theory of multimedia learning (illustrated in Figure 2) was proposed to explain multimedia learning and inform multimedia designs (Mayer, 1997, 2001, 2003, 2005a, 2009; Mayer & Moreno, 2003). The theory was based on three assumptions:

1. A multimedia message is processed separately in an auditory and a visual channel;

2. Each of the auditory and visual channels can only process a limited amount of information at any given moment;
3. Selection, organization, and integration of new information occur when multimedia messages are processed.

Multimedia learning is illustrated in Figure 2. Specifically, there are three memory stores:

1. Sensory memory briefly holds graphic, textual, and auditory inputs;
2. Working memory selects information from sensory memory for processing and integration;
3. Long-term memory stores knowledge base.

When new information is presented verbally and visually by multimedia, it is first received in learner's sensory memory. In sensory memory, learner's ears receive auditory input and eyes receive graphic and textual inputs. Then, the learner selects relevant sounds for processing in auditory working memory and relevant images for processing in visual working memory. Next, the learner organizes selected sounds into a verbal model and images into a pictorial model. Last, the learner integrates verbal and pictorial models with each other and with prior knowledge.

According to Mayer (2009), three demands are involved during multimedia learning: extraneous processing, essential processing, and generative processing.

1. Extraneous processing is the cognitive processing that is extraneous to the goal of the learning and should be minimized during the design of multimedia instruction. Mayer (2009) contends that the extra processing brought about by extraneous (but sometimes interesting) graphics in multimedia instruction can overload students' visual channel, and therefore should be removed.

2. Essential processing is the primary cognitive processing related to the goal of learning. It involves selecting and some organizing of information, which leads to rote learning or good retention of knowledge (Mayer & Moreno, 2003).
3. Generative processing is the deep cognitive processing related to the goal of the instruction. It involves both organizing and integrating the new information, which leads to meaningful learning. According to Mayer and Moreno (2003), meaningful learning involves both good knowledge retention and good knowledge transfer.

Therefore, multimedia lessons should be manipulated to "reduce extraneous processing, manage essential processing, and foster generative processing" (Mayer 2009, p. 57) so that most cognitive resources can be reserved for effective learning.

The Redundancy Principle and the Modality Principle

Building on a number of empirical studies on university students' learning via multimedia (Clark & Mayer, 2011), Mayer (2009) proposed 12 multimedia principles to direct the pedagogy of multimedia instruction. They are the multimedia principle, the coherence principle, the pre-training principle, the spatial contiguity principle, the temporal contiguity principle, the segmenting principle, the modality principle, the redundancy principle, the signaling principle, the voice principle, the image principle, and the personalization principle. Of the principles, the modality and redundancy principles pertain to multimedia input modes.

The redundancy principle. According to the redundancy principle (Moreno & Mayer, 2002), in a multimedia lesson, people learn more effectively from graphics + audio than from graphics + audio + on-screen text. Because of the competition of cognitive resources in the visual channel (Moreno & Mayer, 1999; Tindall-Ford, Chandler, & Sweller, 1997), processing both graphics and text at the same time can lead to an overload of the visual channel, and

therefore impede learning. For example, in a geography classroom where students are watching a video clip to learn about El Niño, there is no need for the instructor to turn on the closed captions (i.e. verbatim on-screen text) while the clip is being played. Processing both graphics and text could overload students' visual channel and distract them from effective learning.

Later, Mayer and Johnson (2008) went on to extend the redundancy principle by examining redundant keywords from the narration next to the corresponding portion of the diagram. They modified the original redundancy principle, because they found that redundant keywords led to superior performance in the subsequent knowledge retention test. Based on the cognitive theory of multimedia learning, keywords direct learners' attention to target content knowledge, and placing keywords next to their corresponding graphics aids the organization and integration of the target content knowledge, and therefore facilitates generative processing and promotes learning. Mayer (2014c) also believed that "the modality principle and the redundancy principle can be modified by including a minimal number of words on the screen mainly, to help highlight the main points and to concretize technical or unfamiliar terms." (p. 68). For example, in a video clip about bees, students can benefit if the text of "beehive" is placed adjacent to the graphics of a beehive on the clip. In this case, the added short text does not lead to a visual channel overload. Instead, it facilitates the integration and organization of new information for more effective generative processing, which promotes learning.

The modality principle. According to the modality principle (Moreno & Mayer, 1999), students learn more effectively from a multimedia message delivered in graphics + audio than in graphics + on-screen text. The principle also stems from the assumption that competition for cognitive resources in the visual channel from the simultaneous presentation of graphics and text can easily lead to a cognitive overload of the visual channel (Moreno & Mayer, 1999). Such an

overload is not conducive to the integration and organization of new information, and therefore, generative processing for effective learning is undermined. For example, an animation with auditory explanation can be more effective for learning the concept of photosynthesis than the same animation with explanatory text on the bottom of the screen.

Of the three input modes involved in the redundancy and modality principles as illustrated in Table 1 (i.e. graphics + text, graphics + audio, and graphics + audio + text), the input mode of graphics + audio should be the most conducive to learning. Since text in the other two modes of graphics + audio + text, and graphics + text could lead to a visual channel overload, generative processing for effective learning is undermined.

Table 1

Comparisons of Input Modes in the Modality and Redundancy Principles

Input Modes	**The Modality Principle**	**The Redundancy Principle**
Graphics + Text	✓	
Graphics + Audio	✓✓	✓✓
Graphics + Audio + Text		✓

Note: ✓ indicates the input mode involved; ✓✓ indicates the more effective input mode.

Conditions for the two principles. According to Mayer (2014b), there is a need to "delineate the boundary conditions under which the principles apply" (p.67). He argued that the modality principle might not apply "when the verbal material contains technical terms, is in the learner's second language, or is presented in segments that are too large to be held in the learner's working memory" (Mayer, 2014b, p. 65) and the redundancy principle could become invalid "when the learners are experienced, the on-screen text is short, or the material lacks graphics" (Mayer, 2014b, p. 63).

Additionally, Clark and Mayer (2011) discussed the conditions for the redundancy and modality principles. First, the modality principle applies on the condition when graphics and their corresponding verbal commentaries match and are presented at the same time. Second, when the multimedia presentation is complex and is in a rapid continuous fashion, text will easily lead to visual channel overload. Third, the redundancy principle applies when verbal information and graphics are shown together at a fast-continuous pace. Fourth, when terminology is present, written words should be provided, because they "aid cognitive processing by directing the learner's attention" (Clark & Mayer, 2011, p. 144). Clark and Mayer (2011) also argue that when the instructional language is not learners' native language, they might have difficulty processing the spoken words of the instructional language, and therefore on-screen text could support understanding and should be available.

Although both Mayer (2014b) and Clark and Mayer (2011) believed that language learners might benefit from on-screen text, there is little empirical evidence to support such an assumption. This study filled this gap by comparing all the three input modes to examine if text indeed supports ESL students' learning.

Verbal message and its graphics in the two principles. At the core of the redundancy and modality principles (Mayer, 2009) is the assumption that simultaneous presence of graphics and text can cause overload in a learner's visual channel. In other words, the modality and redundancy principles only are relevant when both verbal and non-verbal components appear in the multimedia presentation. This echoes the multimedia principle suggesting the need to add graphics to instruction for better learning outcomes (Mayer, 2009). When there are no graphics or when the multimedia principle is not followed, the modality and redundancy principles do not apply (Mayer, 2014b).

Additionally, for the modality and redundancy principles, Mayer (2009) failed to discuss a common condition that verbal messages (textual and auditory) and their corresponding graphics need to closely match. When a verbal message (textual and auditory) and its corresponding graphics do not completely match, there is a possibility that text does not cause visual channel overload. For example, when an instructional video is showing a biology professor sitting in a chair talking about photosynthesis, the graphics (i.e. the talking professor) does not match the verbal message (i.e. the process of photosynthesis). In this case, students might give little attention to the graphics, because the graphics provide little support to the learning of the content knowledge. Graphics and text might not overload the visual channel. Therefore, only when graphics are closely related to audio or text, does the modality and redundancy principles apply. This condition also echoes the coherence principle (Mayer, 2009) that stresses the importance of eliminating extraneous information not conducive to learning. Therefore, in this study, to examine the applicability of the modality and redundancy principles to ESL students, graphics are controlled to match their verbal messages in the multimedia lesson, which is elaborated in Chapter 3.

Other multimedia learning principles and the two principles. To examine the applicability of the modality and the redundancy principles for ESL students' learning, variables that might influence the applicability of the two principles should be removed first. Among the variables are certain multimedia learning principles. During the design of an effective multimedia lesson, multiple multimedia learning principles (Mayer, 2009) are usually taken into consideration. At the core of the modality and redundancy principles is the assumption that simultaneous presence of graphics and text overloads visual channel and thus undermines learning. Therefore, the multimedia principles that have an impact on the assumption might

become conditions of the modality and redundancy principles. In addition to the above-discussed multimedia and coherence principles, some other multimedia learning principles individually can impact the applicability of the modality principle, the redundancy principle, or both for native English-speaking students.

The segmenting principle and the two principles. The segmenting principle states that people learn better if a multimedia lesson is segmented based on his or her own pace (Mayer, 2009). When the learner controls the pace of a multimedia lesson, redundant text is unlikely to add to his or her processing demands to cause visual channel overload (Clark & Mayer, 2011) because learner-controlled pace provides the learner with ample time to fully process all incoming inputs. In such a case, the redundancy principle stating that graphics + audio + text is less conducive to learning than graphics + audio might not apply. Similarly, if the learner can control the pacing of the multimedia lesson, then the modality principle can become less salient (Clark & Mayer, 2011, Moreno, 2006), which is supported by an empirical study of Cheon, Crooks and Chung (2014) who suggested that active pause controlled by the learner had counteracted the modality principle. In this study, to investigate the applicability of the modality and redundancy principles for ESL students, the segmenting principle was not followed in the multimedia lesson to ensure that the modality and redundancy principles were not counteracted due to its presence.

The contiguity principles and the two principles. Audio and text in the modality and redundancy principles need to be simultaneously presented near their corresponding graphics to ensure effective learning (Clark & Mayer, 2011), which echoes the spatial contiguity principle (Mayer, 2009) suggesting text being placed next to its corresponding graphics, and the temporal contiguity principle (Mayer, 2009) suggesting audio being presented at the same time with its

corresponding graphics. For example, when audio is not in line with its corresponding graphics, learners need to utilize additional cognitive resources to make sense of the incoming auditory input. This is not conducive to learning and should be avoided. Therefore, both the spatial contiguity principle and the temporal contiguity principle were followed in the multimedia lesson of this study.

The image principle and the two principles. According to the image principle (Mayer, 2009), a talking head in a multimedia lesson is of little help to learning and should be removed. For the modality and redundancy principles, graphics in a multimedia lesson need to visually demonstrate their corresponding verbal message. When graphics, such as a talking head, is unsupportive to the target verbal message and do not facilitate learning, learners might adjust their viewing and neglect the graphics. In such circumstances, graphics of a talking head and text do not overload the visual channel. Therefore, when the image principle is not followed, the modality and redundancy principles do not apply, which is supported by empirical evidence. Huang, Shadiev and Hwang (2016) and Shadiev, Huang and Hwang (2017) used the same talking head in their multimedia lesson about photography. They both observed that redundant text promoted learning and the redundancy principle did not apply. Therefore, to examine the applicability of the modality and redundancy principles to ESL students' learning, the condition that the graphics visually demonstrated the verbal message was followed in the multimedia lesson of this study.

The combination of multimedia learning principles and the two principles. As discussed above, each of the multimedia principles— the coherence principle, the spatial contiguity principle, the temporal contiguity principle, segmenting principle, and the image principle—has an impact on whether text and graphics together would overload the visual channel, and hence

the applicability of the modality and redundancy principles. The discussion about the modality and redundancy principles is built on an effective multimedia lesson conducive to learning. Ineffective multimedia lessons, such as one with asynchronous words and graphics that contradict the temporal contiguity principle (Mayer, 2009), would not be used for instruction in the first place, let alone for comparisons of different input modes for the modality and redundancy principles. Therefore, certain multimedia learning principles simultaneously apply in an effective multimedia lesson to ensure its conduciveness to learning.

Unlike other studies on multimedia input modes for ESL students' learning (e.g. Mayer et al., 2014), the initial Study 1 examined the modality and redundancy principles for ESL students when the multimedia lesson was controlled by following a combination of the multimedia principle, the coherence principle, the spatial contiguity principle, the temporal contiguity principle, the pre-training principle, the signaling principle, the voice principle. The multimedia lesson in this study was controlled to also follow the personalization principle, the image principle, and the modified redundancy principle, none of which were explicitly explained in Study 1. The control of the multimedia lesson of this study is elaborated in Chapter 3.

Uncontrolled multimedia lessons, such as one that follows the segmenting principle (Mayer, 2009) or is learner-controlled, do not guarantee the conditions for the modality and redundancy principles, and the findings about the applicability of the modality and redundancy principles for ESL students could be due to the controlled multimedia lesson, which is a threat to internal validity. Therefore, to investigate the applicability of the modality and redundancy principles for ESL students, a combination of multimedia learning principles was followed to control the multimedia lesson in this study, which ensures that the multimedia lesson was conducive to learning, and the conditions for the modality and redundancy principles were met.

ESL students and the cognitive theory of multimedia learning. Based on empirical studies on native English-speaking students, the cognitive theory of multimedia learning and the modality and redundancy principles (Mayer, 2009) were proposed to direct designs of effective multimedia instruction for native English-speaking students. ESL students were excluded from the discussion. According to the cognitive theory of multimedia learning (Mayer & Moreno, 2003), text simultaneously presented with graphics leads to an overload in the visual channel and hinders learning, which is also the foundation for the modality and redundancy principles.

However, when a multimedia lesson is presented in English text and audio, ESL students' less proficient English reading and listening skills (Nelson, Balass & Perfetti, 2005; Hirai, 1999; Wong, 2001) might compromise the modality principle and redundancy principles that include English text and audio. The applicability of the two principles for ESL students has not been directly examined until Study 1 (that will be elaborated later).

The Cognitive Load Theory

The cognitive load theory (Sweller, 2014) consists of a few assumptions:

1. There are two kinds of memory: working memory and long-term memory;
2. New information must be stored in working memory for some time before it is processed and becomes integrated into permanent memory;
3. Working memory is short-term and limited in its cognitive capacity in holding the amount of information at once.

Therefore, when a learning task requires much working memory capacity or too much information is presented at the same time, working memory can become overloaded and much of the new information will be lost (Sweller, 2014).

Types of cognitive load. According to the cognitive load theory (Sweller, 2014), there are three kinds of cognitive load: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load.

Intrinsic cognitive load is related to the inherent characteristics or difficulty of the target content knowledge (e.g. calculating 9+3) and it cannot be altered; extraneous cognitive load is irrelevant for learning and results from poor instructional designs (e.g. the redundant verbatim on-screen text in a PowerPoint presentation) that consume unnecessary cognitive resources; germane cognitive load is directly related to schema construction and contributes to learning (e.g. relating the new information to prior knowledge), and therefore, should be maximized with instructional intervention.

An effective instructional design needs to address the limitations of working memory by minimizing the extraneous cognitive processing that is not conducive to learning so that more cognitive resources in the working memory can be allocated to germane cognitive processing that leads to learning (Sweller, 2014).

ESL students and the cognitive load theory. The cognitive load theory also focuses on the content knowledge learning of native English-speaking students, while ESL students were not considered in the discussion. ESL students who are learning new knowledge via English have not mastered the automatic processing of English auditory input comparable to their native English-speaking counterparts. ESL students "may have difficulty in segmenting the incoming flow of sounds into discreet words and therefore must allocate their limited cognitive processing resources to consciously perceiving each word" (Mayer et al., 2014, p. 653) and they" have to spend cognitive resources on decoding unfamiliar sounds that move by rapidly." (Lee & Mayer, 2018, p. 650). At this time, a big portion of cognitive resources in the working memory is

allocated to identifying and decoding English auditory input, and limited resources are available for comprehension and other deeper processing tasks, which negatively impacts ESL students' content knowledge learning.

Added verbatim text can reduce the intrinsic cognitive load resulting from decoding incoming English auditory input (Meskill, 1996), which allows more cognitive resources to become available for learning. According to Lee and Mayer (2018), for university students whose first language was not English, "presenting multimedia lessons as video and printed text is a way to reduce cognitive load and support deeper learning." (p. 650). For ESL students' content knowledge learning, to provide redundant verbatim on-screen text might "preserve word availability (as spoken words are transitory), making it easier for students to encode the words" (Mayer et al., 2014, p. 653). Similarly, verbatim text "helps learners segment what might otherwise be an incomprehensible stream of speech" (Winke, Gass & Sydorenko, 2010, p. 65). Similarly, Jung (1990) suggested that redundant text could "diminish the decoding load placed upon the learner by the unrefined audio signal of authentic speech" (p. 208-209). Additionally, to discuss the amount of cognitive demands brought about by audio and its verbatim text, Meskill (1996) acknowledged that the cognitive demands for second language learners to decode the auditory input were strong, and argued that the availability of text in ESL students' multimedia instruction could enable students to "attend to overall meaning derivable from multiple representations of input"(p.185), and "support comprehension of the message as opposed to drawing attention to its constituent parts" (p.185). Lambert (1986) argued that on-screen text in multimedia instruction could serve as advance organizers that scaffold and support auditory input to reduce working memory load.

To summarize, based on the cognitive load theory, verbatim on-screen text of auditory input reduces ESL students' identifying and decoding load, which promotes their learning.

Text in the Modality and Redundancy Principles for ESL Students

As explained above, the modality and redundancy principles did not focus on ESL students. The roles of on-screen text in the two principles are different for native English-speaking students and ESL students as illustrated in Table 2. On the one hand, the modality and redundancy principles imply the detrimental role of added text for both native English speakers and ESL students, because it leads to a visual channel overload. On the other hand, added text can promote ESL students' learning by reducing the identifying and decoding load of English audio (Mayer et al., 2014). Although Clark and Mayer (2011) also believed that language learners could benefit from added text, the facilitative role of text was not empirically verified, nor was the applicability of the modality and redundancy principles for ESL students.

Table 2

Roles of Text for the Two Populations Based on the Two Theories

Text	The Cognitive Theory of Multimedia Learning	The Cognitive Load Theory
Native English Speakers	Text and graphics lead to a visual channel overload.	Added text leads to a working memory overload.
ESL Students	Text and graphics lead to a visual channel overload.	Text reduces the identifying and decoding load of L2 auditory input.

Therefore, this study attempted to fill this gap by comparing ESL students' learning from the three input modes (i.e. graphics + audio, graphics + text, graphics + audio + text) in the modality and redundancy principles in order to examine the applicability of the two principles for ESL students.

The Media Debate

The great "media debate" has centered around media's influence on learning. Clark (1983,1994) and Kozma (1991,1994) are two contending parties in the debate.

Based on the results of a meta-analysis investigating the influence of media on learning, Clark (1983) claimed that media did not influence learning under any circumstances, and he argued that it was instructional methods rather than media that had the most influence on learning. Clark (1999) utilized a metaphor that a truck that delivered food would never influence the nutrition of food in the truck to describe the relationship between media and learning.

Komza (1991) believed that media was more than a vehicle for delivery. Kozma (1991) attempted to negate Clark's (1983) claim by proposing a new theory outlining the synergy of content, media, and interaction between learner and environment. He argued that media varied by technology, symbol systems, and processing capabilities, and certain medium attributes were more suitable than others for a given task (Kozma, 1991). However, Clark (1994) proposed the replaceability test implying that there were always a variety of attributes that could produce the same learning goal.

Kozma (1994) contended that "if we can find a relationship between media and learning then we will be able to see how technology influences learning"(p. 8). This study examined if the three multimedia input modes of graphics + audio, graphics + text, and graphics + audio + text would lead to different learning outcomes for ESL students, which was an opportunity to test if there was a relationship between media and learning. Findings of this study could provide new evidence to the debate.

Previous Studies

What defines ESL students is their insufficient English proficiencies. Therefore, empirical studies on ESL students inherently focus on their L2 English language learning rather

than content knowledge learning. In this section, empirical studies that examined input modes and their influences on adult ESL students' comprehension and vocabulary acquisition were reviewed and discussed, so were several studies examining input modes and their efficacy for ESL students' content knowledge learning. The input modes include any one featuring audio, graphics and text. Studies on language learning for ESL students mainly focus on vocabulary acquisition and comprehension.

Graphics and ESL Students' Language Learning

According to Mayer (2009), presence of graphics is a defining characteristic of multimedia instruction. The multimedia principle (Mayer, 2009) states that people (i.e. native English speakers) learn better from graphics and texts than from texts alone, supporting adding graphics for better learning. Compared with native English speakers, ESL students have incomparable L2 English listening and reading proficiencies to access verbal messages; however, they are equally able to access messages through graphics. Empirical evidence supports adding graphics for better language learning of adult ESL students (Akbulut, 2007; Granoff &Whiting, 2010; Rivas, 2011; Shin, 1998; Smidt & Hegelheimer, 2004; Tsou, Wang, & Li, 2002; Wagner, 2010).

Graphics and ESL students' vocabulary acquisition. Non-verbal input of graphics including pictures, illustrations, and animations etc. was found to be able to complement verbal input for better learning of L2 target words (Akbulut, 2007; Rivas, 2011; Tsou, Wang, & Li, 2002). Only Akbulut (2007) targeted adult rather than young ESL students. He examined three types of interactive multimedia annotation and their efficacy to advance university-level ESL students' vocabulary learning. The three types of multimedia annotations were: multimedia textual definitions (i.e. text), multimedia textual definitions with pictures (i.e. text + graphics),

and multimedia textual definitions with videos (i.e. text + graphics + audio). Graphics were found to have significantly improved ESL students' performances in both immediate and delayed posttests (Akbulut, 2007).

Graphics and ESL students' comprehension. Empirical studies (Granoff &Whiting, 2010; Guichon & McLornan, 2008; Shin, 1998; Smidt & Hegelheimer, 2004; Wagner, 2010) indicated that added graphics aided in ESL students' comprehension.

Granoff and Whiting (2010), Shin (1998), and Wagner (2010) all examined the influences of added graphics on comprehension by comparing the input mode of graphics + audio with that of audio alone. In the early study of Shin (1998), 83 university ESL students were randomly assigned to either a video (i.e. graphics + audio) group or an audio only group. A pretest of multiple-choice questions assessed listening proficiency before the experiment. Posttest results showed that ESL students in the video group outscored those in the audio only group in listening comprehension. The author suggested that added graphics in the video led to superior listening comprehension. However, an issue with the design of this study was that the two versions of the post-test were not identical. The posttest for the audio-only group was modified to eliminate long silences and to correct grammatical errors, while the posttest for the video group was not. Even though the posttest for the video group was harder, adult ESL students in the video group still outperformed those in the audio-only group in comprehension.

Similarly, Wagner (2010) examined the effect of graphics on university-level ESL students' listening test performance. Two hundred and two university ESL students from various ethnic backgrounds participated in the study. The randomly assigned experimental group was tested with a video (i.e. graphics + audio), while the control group was tested with an audio. The listening posttest scores across two genres of test materials indicated that ESL students in the

graphics-enriched group outscored those in the control group in overall scores and in scores of both genres, which suggested that the added graphics led to superior listening comprehension. Additionally, Granoff and Whiting (2010) investigated university-level ESL students' comprehension of a story from a video (i.e. graphics + audio), and from auditory input alone. They found that added graphics did facilitate ESL students' comprehension of the story.

Smidt and Hegelheimer (2004) incorporated three input modes in their discussion. The researchers compared university-level ESL students' listening comprehension from three conditions of a talking head, a talking head with keywords, and a talking head with an illustration slideshow (i.e. added graphics). Illustrations were found to have yielded statistically significant gains in enhancing listening comprehension. This finding supported a more complex study of Guichon and McLornan (2008) who compared native French-speaking university students' comprehension of a BBC recording from four input modes of audio alone, audio + graphics, audio + graphics + L2 (i.e. second language) text, audio + graphics + L1 (i.e. first language) text. Results of a production test also indicated that added graphics significantly increased students' comprehension.

In summary, added graphics seemed to have had a positive impact on ESL students' language learning. In other words, the multimedia principle that added graphics facilitates learning (Mayer, 2009) seemed to also apply to ESL students' language learning. However, there is little empirical evidence about the applicability of the multimedia principle for ESL students' content knowledge learning.

Text and ESL Students' Language Learning

The input mode of graphics + audio + text has been widely compared with the mode of graphics + audio to investigate the efficacy of added text or caption on the development of L2

English proficiency. As a textual form of verbal input, caption (i.e. verbatim on-screen text of auditory input) is traditionally used to support auditory input for the hearing impaired. Currently, caption is increasingly used for the learning and instruction of ESL students (Rodgers & Webb, 2016). However, the benefits of caption for ESL students' language learning are less conclusive.

Caption and ESL students' vocabulary acquisition. Empirical studies focused on examining ESL students' incidental vocabulary learning from captioned (i.e. graphics + audio + text) and non-captioned (i.e. graphics + audio) video clips. Caption was found to have a facilitative impact on ESL students' vocabulary acquisition (Aldera & Mohsen, 2013; Bird & Williams, 2002; Markham, 1999). Markham (1999) explored the impact of captioned videotapes on advanced adult ESL students' listening word recognition. One hundred and eighteen ESL students were randomly assigned to a caption group and a non-caption group. The instructional material consisted of episodes from two separate educational television programs. The results of listening tests revealed that caption greatly enhanced students' performance in identifying the words both on the videotapes and on the tests. Additionally, Bird and Williams (2002) conducted one of the most robust and fundamental experiments along this line. They investigated if a bimodal presentation of L2 English novel words would affect the learning of the words. In their study, new vocabulary was presented to advanced ESL students under three conditions: (a) text + audio, (b) text alone, and (c) audio alone. Recognition of spoken words and memorization were assessed. Posttest results indicated that the mode of text + audio led to superior recognition memory of spoken words and non-words compared with the other two input modes. The researchers argued that the bimodal input mode of text + audio promoted word learning, because the processing of auditory input was facilitated by text. In a recent study by Aldera and Mohsen (2013), native Arab-speaking ESL students watched an animation under one of three conditions:

(a) graphics + audio + text + keyword annotation, (b) graphics + audio + text, and (c) graphics + audio. After four weeks, the students completed vocabulary recognition and vocabulary production tests. The ESL students in the two groups with added text were found to have significantly outperformed those in the graphics + audio group in both vocabulary tests.

Additionally, Koolstra and Beentjes (1999), and Neuman and Koskinen (1992) both focused on young ESL students and reported that captioned videos led to gains in L2 word recognition. Montero Perez, Van Den Noortgate and Desmet (2013) conducted a meta-analysis and found that there was a large effect of caption on L2 vocabulary learning.

Caption and ESL students' comprehension. Different from the general consensus on the facilitative role of caption for ESL students' vocabulary learning, mixed findings about the role of caption for ESL students' comprehension were observed.

Caption promoted ESL students' comprehension. Guichon and McLornan (2008), Hayati and Mohmedi (2011), Huang and Eskey (1999), and Toh, Munassar and Yahaya (2010) all examined ESL students' comprehension of authentic English-speaking videos with and without caption. In all those studies, ESL students watching with caption on performed better in subsequent comprehension tests. Additionally, Chung (1999) compared four types of viewing conditions in terms of their influences on comprehension. Adult Taiwanese students were randomly assigned to four conditions: graphics + audio, graphics + audio + text, graphics + audio + advance organizers, and graphics + audio + text + advance organizers. The presence of both text and advance organizers was found to have led to the best comprehension, and caption was also found to have promoted comprehension.

The reverse redundancy principle that caption promoted rather than impeded comprehension was observed in the above studies. To explain the benefits of caption for ESL

students' comprehension, some researchers (e.g. Bird & Williams, 2002) suggested that caption was likely to promote greater depth of spoken-word processing. Some also argued that caption "helps learners segment what might otherwise be an incomprehensible stream of speech" (Winke et al., 2010, p. 65). Others (e.g. Jung, 1990) believed that caption was able to reduce the cognitive resources used for decoding English auditory input for better comprehension.

Other findings about caption and ESL students' comprehension. Due to different research questions, different methodologies, and mixed findings, it remains inconclusive if caption promotes ESL students' comprehension.

Rodgers and Webb (2016) examined university-level Japanese ESL students' comprehension of long and short episodes of an authentic TV program with or without the availability of caption. Comprehension tests indicated that captioning seemed to have only led to superior comprehension of difficult episodes, and comprehension of short episodes did not seem to have been affected by the presence of caption. Additionally, Livesidge (2000) also examined university-level Japanese ESL students' comprehension of two different movie clips (i.e. *Graduate* and *Airplane*). Caption seemed to be beneficial to understanding only in some comprehension tests but not others. According to these studies, multimedia materials can be a possible variable in investigating captioning for ESL students' comprehension.

Aldera and Mohsen (2013), and İnceçay and Koçoğlu (2016) observed that the added caption impeded listening comprehension. Aldera and Mohsen (2013) found that caption facilitated vocabulary acquisition. However, comprehension test results indicated that the animation-alone (i.e. graphics + audio) group significantly outscored the caption (i.e. graphics + audio + text) group in listening comprehension. The researchers argued that caption might have created extra cognitive load and therefore hindered comprehension. Similarly, İnceçay and

Koçoğlu (2016) compared university-level Turkish ESL students' listening comprehension scores from four input modes: audio, graphics + audio, graphics + audio + text, and audio + PowerPoint presentation. Listening comprehension test results revealed that the mode featuring text led to the lowest test scores, and the mode was also reported to have caused ESL students most confusion and anxiety in comprehension.

A majority of studies (Aldera & Mohsen, 2013; Bird & Williams, 2002; Chung, 1999; Huang & Eskey, 1999; Hwang & Huang, 2011; Liversidge, 2000; Neuman & Koskinen, 1992; Toh et al., 2010) assessed comprehension right after the treatment. Studies addressing long-term development of literacy skills are of greater empirical significance (Markham, 1999). Hwang and Huang (2011) examined development of long-term comprehension skills. Two groups (i.e. 40 for each group) of freshmen in a university in Taiwan watched two versions of instructional videos with and without caption one hour every two weeks. After ten weeks, reading comprehension test (as part of General English Proficiency Test developed in Taiwan) was administered in order to determine if captioning could promote learners' reading comprehension achievement. The results showed no significant gains in reading comprehension competence for the students who watched captioned videos. The researchers suggested that the treatment was only ten weeks long, and it might be too short for any statistically significant gains to manifest.

Limitations of studies addressing caption and ESL students' language learning. Studies on caption and ESL students' language learning suffered from the following common validity issues that informed this study.

First, as Liversidge (2000) suggested, "very few studies have provided a way by which material could be assessed for suitability for in-class or self-study activities" (p. 22-23). Although there is evidence suggesting that multimedia lessons could be a variable (Livesidge,

2000; Rodgers & Webb, 2016), none of the multimedia materials in the discussed studies were controlled. For a multimedia lesson, how text was used by ESL students could be related to how useful the graphics were: the more useful the graphics in facilitating learning, the less ESL students needed verbatim text to support their learning. For example, Toh et al. (2010) used an instructional animation to examine the impact of captioning for ESL students' comprehension. In their animation, the graphics consisting of a few cartoon pictures conveyed little supportive information to comprehension. In this case, ESL students might need to adjust their viewing by neglecting the graphics and resorting more to caption to support comprehension, which could account for the finding that captioning facilitated comprehension. However, if graphics, such as the ones in the multimedia lesson of Mayer et al. (2014), were highly supportive to learning, ESL students might adjust their viewing by neglecting L2 English verbal input and focusing on graphics to learn, which could account for the finding of Mayer et al. (2014) that captioning did not promote learning. Therefore, usefulness of graphics can have an impact on how text is used by ESL students, and then learning outcomes from the modes of graphics + audio and graphics + audio + text. If the multimedia material is not controlled, it is hard to come to any rigorous conclusion about the role of caption for ESL students.

Second, certain conditions for vocabulary learning were not addressed in the multimedia materials of the discussed studies. Conditions, such as motivation (Tseng & Schmitt, 2008), the number of exposures or repetitions of target words (Webb, 2007), noticing or conscious attention (Schmidt, 1990), engagement and attention (Butler et al., 2010), and contextualization (Brown, 1993), can have an impact on the outcomes of vocabulary learning. For example, students watching an uninteresting video clip without any learning goal in mind might pay very little attention to the information in the clip, which makes learning difficult to happen. Therefore,

when these conditions were not controlled, input modes alone could not account for the learning outcomes, which was a threat to internal validity.

Additionally, Montero Perez and Desmet (2012) questioned the construct validity of the vocabulary tests used in the studies and argued, "only a small part of vocabulary knowledge measured by the pre and posttests can be attributed to the treatment" (p.156). According to Nation (2001), knowing a word should involve knowing its form, meaning, and use. Different aspects of vocabulary knowledge are evaluated by vocabulary tests of different formats. It is important to determine which aspect of vocabulary knowledge (of meaning, form, or use) should be assessed (Nation, 2001) and what kind of vocabulary knowledge (e.g. receptive vs. productive) is most suitable to measure in a vocabulary test (Joe, Nation, & Newton, 1996). Therefore, it is inaccurate to conclude that added text has a positive effect in learning vocabulary, since no relevant studies addressed all aspects of form, meaning, and use.

Last, among the studies addressing L2 vocabulary acquisition from multimedia, technical or content specific vocabulary (Diamond & Gutlohn, 2006) that is often an important part of content knowledge learning was barely examined. Since the related studies focused on ESL students' language learning rather than content knowledge learning, the highly content-specific technical terms were hardly investigated. According to Hinkel (2005), 20% to 30% of the running vocabulary in content area text is comprised of technical vocabulary that is often formally defined. The meanings or definitions of technical terms are often the focus of vocabulary learning as well as content knowledge learning (Hinkel, 2005). Therefore, it is of empirical significance that ESL students' learning of technical vocabulary from multimedia instruction is examined.

Therefore, to address the above limitations, in this study, the multimedia material was

controlled in terms of the usefulness of graphics. Additionally, technical terms were explicitly taught in the multimedia lesson of this study, and the subsequent vocabulary test on terminology focused on assessing the retention of meaning.

Multimedia Input Modes and ESL Students' Content Knowledge Learning

Since a vast majority of ESL students come to the US to learn content knowledge in specific disciplines rather than merely improve their English skills, it is essential that instructional multimedia made for native English-speaking students also enhance the content knowledge learning of ESL students. Content knowledge learning refers to the learning of new information in a specific content area (e.g. a chemical reaction, and a meteorological explanation). However, since the defining characteristic of ESL students is their lack of comparable English language skills, a great body of research on ESL students (e.g. Bird & Williams, 2002; Chung, 1999; Davey & Parkhill, 2012; Etemadi, 2012; Garza, 1991; Guichon & McLornan, 2008; Hayati & Mohmedi, 2011; Huang & Eskey, 1999; Liversidge, 2000; Markham, 1999; Markham & Peter, 2003; Rowland 2007; Toh, Munassar, & Yahaya, 2010) focuses on the development of their language skills. Empirical studies comparing different multimedia input modes and ESL students' learning of content knowledge are scarce until recently.

Audio and ESL students' content knowledge learning. Kozan, Erçetin and Richardson (2015) examined native Turkish-speaking university students' learning about tornado formation. There were two input modes: audiovisual (i.e. graphics + audio + text) and visual only (i.e. graphics + text). Learning was assessed by immediate retention and transfer tests, as well as delayed retention and transfer tests. Results revealed a significant combined effect of the input mode, time, and working memory capacity on knowledge retention. Specifically, the added audio

benefited those with high working memory more regarding retention of information over time. In addition, only time had a significant effect on ESL students' performance on transfer tests.

Caption and ESL students' content knowledge learning. Caption or added text was found to have promoted content knowledge learning in some studies, but not in others.

Lin, Lee, Wang and Lin (2016) examined university-level Taiwanese ESL students' learning from an instructional video about the human brain with and without the availability of caption and e-notes. Through a 2 (English subtitle/no) x 2 (taking e-notes/no) factorial design, they found that caption decreased ESL students' cognitive load, and subsequently increased their posttest performance. The authors suggested a reverse split-attention effect for ESL learners. The beneficial role of caption for ESL students' content knowledge learning (Lin et al., 2016) was supported by another two similar studies. Huang et al. (2016) and Shadiev et al. (2017) both investigated the impact of verbatim on-screen text generated from speech to text recognition on university-level ESL students' multimedia learning about photography and dating. The graphics in both studies consisted of lecture slides and a talking head. Added verbatim text was found in both studies to have had a facilitative effect on ESL students' learning. It is noteworthy that graphics in their multimedia lesson merely consisted of a talking head, rather than the visual demonstration of the target content knowledge, which contradicted the image principle (Mayer, 2009). As explained in the theoretical framework section about the conditions of the redundancy principle, the image principle conditions the redundancy principle, and should be followed before any discussion about the redundancy principle. Additionally, prior knowledge about the new information to be learned was not assessed in either study, which threatened internal validity.

On the other hand, findings of some studies indicated that caption did not enhance

content knowledge learning. Mayer et al. (2014) compared the input mode of graphics + audio+ text with the mode of graphics + audio to examine if text or caption was able to enhance university ESL students' learning of content knowledge. Seventy-three university-level ESL students were randomly assigned to watch a 9-minute scientific explanation about the formation of a soda fountain with or without caption. Then, they worked on an immediate knowledge retention test and a knowledge transfer test. Results of the tests did not indicate statistically significant difference in the learning outcomes between the experimental groups. The study conducted by Mayer et al. (2014) addressed the gap in the literature by extending the subjects from native English-speaking students to ESL students.

Later, Van der Zee, Admiraal, Paas, Saab and Gisbers (2017) examined the three-way interaction between the availability of caption, learner's English proficiency and the complexity of the instructional video. Adult ESL learners of five English proficiency levels watched four educational videos entitled "The Kidney," "History of Genetics," "The Visual System," and "The Peripheral Nervous System" with or without English subtitles. Captioning was found to have neither a main effect nor any interaction, which confirms the findings of Mayer et al. (2014). However, the complexity of the video and language proficiency did have a significant impact on adult ESL learners' learning.

Limitations of the studies addressing multimedia input modes and ESL student knowledge learning. The discussed studies (Huang et al., 2016; Kozan et al., 2015; Lin et al., 2016; Mayer et al., 2014; Shadiev et al., 2017; Van der Zee et al., 2017) were limited by two common validity threats. First, language use in the instruments was not controlled. The instruments were designed to evaluate ESL students' prior content knowledge and their content knowledge learning. However, ESL students' lack of English proficiency and cultural knowledge

might deny them access to the meanings of the test questions, which was a threat to construct validity. For example, in the prior knowledge survey (Mayer et al., 2014), ESL students might not know the culturally- contextualized term of *Mentos and Diet Coke fountain*, but they might have already understood the mechanism of the term. To ensure that the instruments accurately assess learning, the language use in the instruments should have been controlled. Therefore, in this study, L1 translations of the questions in the instruments were provided, and participants were allowed to answer questions in their L1s.

Second, criteria of the multimedia material selection were not discussed. Van der Zee et al. (2017) argued, “though the current study used four different videos, each with four different versions, this is not sufficient to be able to generalize to all kinds of educational videos” (p. 27). A lack of multimedia selection criteria leads to a threat to internal validity. On the one hand, use of a single uncontrolled multimedia material to investigate different input modes' efficacy on learning threatens construct validity, because a single case scenario without following certain principles cannot adequately exemplify the domain of the construct (Bellini & Rumrill, 2009). Therefore, a multimedia lesson should be controlled to ensure that it is representative of different types of multimedia lessons. On the other hand, if the multimedia material is uncontrolled, it is hard to attribute the different learning outcomes solely to different input modes. Mayer et al. (2014) also acknowledged that their multimedia material (i.e. MythBusters' episode ‘Diet Coke and Mentos') was not from an ecological setting, and therefore might not be conducive to learning. They also suggested “future research is needed to replicate our findings with other materials” (Mayer et al., 2014, p. 659). Not all multimedia are suitable for instruction, and only those that follow multimedia principles to facilitate learning should be used for instructional

purposes. Therefore, it is important that a multimedia lesson is controlled to ensure that it is conducive to learning in the first place.

Therefore, to address the two limitations, the multimedia material in this study was controlled to be both representative of different types of multimedia materials and conducive to learning. Additionally, the instruments of this study were controlled linguistically to accurately assess ESL students' content knowledge learning.

Summary

In this chapter, the theoretical framework, and the related studies were discussed. Below is a summary of how the theoretical framework and the related literature contributed to the understanding of the research questions and informed the methodology to investigate the questions.

The Theoretical Framework

The theoretical framework consists of the cognitive theory of multimedia learning, including the modality and redundancy principles, and the cognitive load theory. Below are how some major findings contributed to the understanding of the research questions and methodology to examine them.

First, the modality and redundancy principles originally did not include ESL students in the discussion. Based on a plethora of empirical studies focusing on native English-speaking students, both cognitive theory of multimedia learning and the 12 multimedia learning principles (Mayer, 2009) were proposed to optimize native English-speaking students' learning outcomes, which justifies the investigation into the applicability of the two principles for ESL students in English-speaking university classrooms.

Second, the modality and redundancy principles consist of the three input modes of graphics + audio, graphics + text, and graphics + audio + text. Each principle was proposed based on the comparison of two input modes in terms of learning outcomes. Therefore, to examine the applicability of the both modality and redundancy principles for ESL students at the same time, all the three input modes could be compared with each other, which informed the design of this study.

Third, both principles imply the detrimental role of verbatim on-screen text for native English-speaking students' learning, because simultaneous presence of text and graphics would lead to a visual channel overload, which undermines learning. However, certain multimedia learning principles could compromise such a visual channel overload, and subsequently become the conditions for the modality and redundancy principles, which informed the control of the multimedia lesson in this study.

Last, based on the cognitive load theory, text could promote learning by reducing the cognitive load brought about by decoding L2 English auditory input, which contradicts the detrimental role of text in the modality and redundancy principle. Mayer (2014b) and Clark and Mayer (2011) also believed that language learners might benefit from on-screen text. However, there is little empirical evidence supporting that text facilitates ESL students' learning, and subsequently the modality and redundancy principles for ESL students do not apply. This study filled this gap by directly examining the role of text, as well as the applicability of the two principles for ESL students.

The Literature Review

Based on the reviewed literature, empirical studies on input modes and ESL students focused on their development of language skills, rather than content knowledge learning.

Additionally, a majority of studies examined the role of redundant text or caption in the input mode of graphics + audio + text, and studies seemed to indicate that caption could have a positive effect in advancing adult ESL students' developments of literacy skills (e.g. Bird & Williams, 2002; Chung, 1999; Davey & Parkhill, 2012; Etemadi, 2012; Garza, 1991; Guichon & McLornan, 2008; Hayati & Mohmedi, 2011; Huang & Eskey, 1999; Liversidge, 2000; Markham, 1999; Markham & Peter, 2003; Rowland 2007; Toh, Munassar, & Yahaya, 2010), and promoting their content knowledge learning (e.g. Huang et al., 2016; Lin et al., 2016; Shadiev et al., 2017). However, the limited number of empirical studies, different methodological constraints of the available studies, great varieties of interventions, and mixed findings do not warrant any definite conclusion to be drawn.

This study responded to five gaps in the related literature dealing with input modes and adult ESL students' multimedia learning. Those gaps contributed to the research questions, as well as informed the methodology to investigate them.

First, to address the gap that the input mode of graphics + text has not been compared with other input modes, this study involved the three input modes of graphics + audio, graphics + text, and graphics + audio + text to determine the most facilitative input mode and explore the role of on-screen text for ESL students' learning. Additionally, the applicability of both the modality and redundancy principles to ESL students' learning could be investigated when the three input modes were compared with each other.

Second, this study addressed the gap that there had been few empirical studies on multimedia input modes and their influences on ESL students' content knowledge learning, rather than their developments of literacy and language skills. The posttest of this study assessed both content knowledge and language learning. Additionally, in the posttest of this study, there

was an assessment on ESL students' learning of highly content-specific technical terms, rather than non-technical terms as in most of the previous studies (e.g. Aldera & Mohsen, 2013; Bird & Williams, 2002; Markham, 1999).

Next, to address the common validity issue of lacking control of the multimedia material, the multimedia material in this study was controlled to ensure that it was both conducive to learning and representative of different types of multimedia materials.

Last, to address the common validity issue of lacking language control of the instruments, L1 translations of keywords and question items were added on the instruments of this study was allowed for answering questions.

CHAPTER 3

METHODLOGY

This chapter addressed the design, participants, instruments, the instructional material, the procedure, as well as scoring and data analysis, all of which served to explore the three research questions. I first described the methodology of the initial study 1, then elaborated the adjustments made in the methodology for study 2.

The research questions are:

1. Which of the three input modes is the most facilitative for ESL students' learning?

2. Does the modality principle apply to ESL students' multimedia learning?

3. Does the redundancy principle apply to ESL students' multimedia learning?

Study 1

The initial Study 1 (Liu et al., 2018) attempted to examine the research questions. It extended the studies that examined input modes and ESL students' content knowledge learning (Huang et al., 2016; Kozan et al., 2015; Mayer et al., 2014; Shadiev et al., 2017) by involving all the three input modes of graphics + text, graphics + audio, and graphics + audio + text in the modality and redundancy principles in the discussion. Additionally, to address the common validity threats in those studies (Huang et al., 2016; Kozan et al., 2015; Mayer et al., 2014; Shadiev et al., 2017), both the multimedia material and the instruments were controlled in Study 1. To fill the gap in the literature, Study 1 also examined the applicability of both the modality and redundancy principles for ESL students.

Research Design

Three input modes that were addressed in the modality and redundancy principles were tested in the present study, as shown in Table 3. For the first question the learning test results from 1PO, 2PO, and 3PO were compared; for the second question results for the graphics + audio group were compared with those for the graphics + text group; and for the third question results for the graphics + audio group were compared with those for the graphics + audio + text group.

Table 3

Three Types of Input Modes

	Pre-tests (PR)	Post-tests (PO)
Graphics + audio (1)	1PR	1PO
Graphics + text (2)	2PR	2PO
Graphics + audio + text (3)	3PR	3PO

Participants

The participants were 48 first-year undergraduate international students studying at a northeastern research university in the US who were recruited from the English language classes provided by the university for ESL students. Three classes were randomly selected from the seven intermediate-level English classes offered. Students were placed in the ESL classes based on a university-administered English placement test. 24 participants were male and 24 were female, and participants' average age was 19.71 years (*SD* = 1.81). The three classes were randomly assigned to the three conditions: 17 participants were in the graphics + audio group;16 were in the graphics + text group; 15 were in the graphics + audio + text group. Chinese was listed to be the first language of thirty-eight participants (79.2%), Korean was listed by 5, Hindi was listed by 2 and Portuguese was listed by 3, which was a representative sample of the international student demographic profile on campus. Their average TOEFL score was 90.76 (*SD*

= 6.46), their average TOEFL Listening score was 23.49 (*SD* = 3.04), and their average TOEFL Reading score was 24.88 (*SD* = 2.84). The participants had been in the US for an average of 11.29 months (*SD* = 11.29), and they had learned English at school for an average of 9.95 years (*SD* = 3.12).

Instructional Material

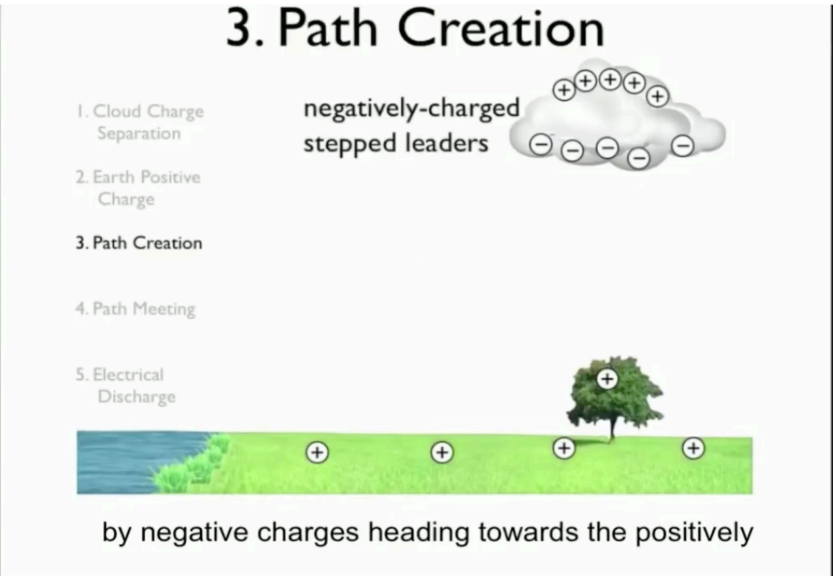

Figure 3. A screenshot from the multimedia lesson in graphics + audio + text.

The multimedia material consisted of a five-minute color video describing lightning. The formation of lightning was an example of the content knowledge to be learned. The instructional video was based on a clip developed by Moulton (2010). With the rapid development of multimedia technology, a color video rather than a black-and-white slideshow used by Moreno and Mayer (2002) is now more likely to be used for instruction. The video began with a five-second red silent screen displaying a numeral (either one, two or three), representing an input mode. The video then briefly introduced the mechanism and five steps of lightning and continued with a detailed description of each step. A video editing software tool CAMTASIA was used for editing including adding verbatim on-screen text (as shown in Figure 3). Text in black Times

New Roman font synchronous with the narration was provided on the bottom of the video. The videos with each type of input mode were uploaded on YouTube.

Table 4

Multimedia Principles Observed in the Design of the Instructional Video in Study 1

Name of Principle	Content of Principle
The coherence principle	Content irrelevant to learning is removed from the animation.
The spatial contiguity principle	On-screen keywords are placed near the corresponding pictures.
The temporal contiguity principle	Graphics and corresponding narrations are presented simultaneously.
The pre-training principle	Keywords and key steps are explained briefly before receiving treatments.
The signaling principle	Keywords and key concepts are highlighted.
The voice principle	People learn better when the narration is spoken in a standard-accented human voice than a machine voice or foreign-accented human voice.

To address the common validity issue that "the lesson itself may have not been as conducive to learning as a well-refined lecture may be" (Mayer et al., 2014, p. 659) in previous studies, the video was controlled to ensure that it was both representative of other types of instructional videos and conducive to content knowledge learning of ESL students. On the one hand, to ensure that the instructional video was conducive to learning, Mayer's multimedia learning principles (2009) in Table 4 were observed during the creation of the instructional video. The video was controlled as follows: based on the coherence principle, there were no irrelevant graphics to the content in the video; based on the spatial contiguity principle, on-screen keywords and the corresponding graphics were placed close to each other; based on the temporal

contiguity principle, narrations were played at the same time as the corresponding graphics; based on the pre-training principle, the key content was explained briefly at the beginning of the video; based on the signaling principle, keywords in the video were highlighted; based on the voice principle, a native speaker of English narrated the video in a clear and moderately-paced fashion. On the other hand, the English language demands were minimized to foster content learning. Five students from the same-level ESL classes as those in which the participants were enrolled were invited to provide feedback on the language used in the video. They all agreed that the narration was clear, moderately paced, and free of unusual accent; the great majority of the vocabulary, except for a few content-specific academic words that were explicitly explained in the video, was easily understandable in both written and auditory forms; and the video was free of unusual social and cultural references.

Instruments

The instruments consisted of a questionnaire, a survey, and a posttest. To ensure the ESL students could easily comprehend all the questions, the English language use was both simple in sentence structure and free of terminology. Additionally, Chinese or Korean translations were added under English instructions. The five ESL participants whose L1 was neither Mandarin nor Korean were asked about their understanding of the questions and were provided oral explanations in simple English when issues arose. The questionnaire elicited participants' demographic information, including their gender, age, first language, and years of English learning at school. To address the issue that no prior knowledge assessment was administered before the experiment to remove those who had already known the topic in previous studies (Huang et al., 2016; Shadiev et al., 2017), this study included a prior knowledge survey. The survey consisted of eight Yes or No questions and one multiple-choice question. It assessed

participants' background knowledge about the field of the content knowledge using a scale from 1 to 5 (1 = *very little,* 5 = *very much*). Mayer et al.'s (2014) survey questions were modified to fit the video content of this study.

The posttest consisted of a knowledge retention test and a vocabulary test. In the retention test, participants were instructed to explain the five steps of lightning and briefly illustrate them with a diagram. Different from the retention tests in Moreno and Mayer (1999) and Mayer et al. (2014) that simply asked the participants to write down all they could remember, the table and the diagram in the retention test of this study were meant to elicit production that could otherwise be undermined by lack of English language proficiency. Additionally, the participants were allowed to use both English and their L1s to answer the questions, which ensured that their lack of English competence would not stand in the way of their production. The learning of Tier 3 technical vocabulary from multimedia instruction has rarely been purposefully examined. There was thus a vocabulary test that asked participants to define some highly content-specific technical terms that constitute an important element of content knowledge. In this context, the definition of a technical term is often the focus of vocabulary learning, with students learning to explain the mechanism of *photosynthesis*, for example. Different formats of vocabulary tests tap different aspects of vocabulary knowledge. The vocabulary test in this study assessed ESL students' receptive knowledge and focused on the retention of meaning.

Procedure

In each of the three ESL classes, the study was briefly explained, and the consent forms were handed out for signing. The researcher then distributed the questionnaires and the surveys. Most of the participants were L1 Chinese speakers, so they were given the questionnaire and the

survey with corresponding Chinese translations, while the five Korean speakers were given the surveys with Korean translations. The researcher asked those participants whose L1 was neither Chinese nor Korean if they had difficulty understanding the questions and where necessary explained and paraphrased the questions in simple English to make sure that they were fully understood. A randomly assigned number (one, two or three) was printed on the upper left corner of each questionnaire. After all participants had completed their questionnaires and surveys, they were directed to turn on their laptops and access a thirty-second sample animation on YouTube. The animation consisted of a few colored pictures of lightning and a synchronous narration that said, "Have you ever wondered how lightning really works? And this is what we will learn today." The researcher gave each participant a pair of headphones and asked them to adjust the volume of the headphones when watching the sample animation. The participants were then directed to access the instructional video on YouTube corresponding to the randomly assigned number on their questionnaire. The researcher repeatedly stressed that participants should stop once they located their assigned video. After making sure that each participant had identified the correct video and stopped there, the researcher explained that the participants were to watch a short instructional video about the formation of lightning and encouraged them to give it their full attention. The researcher also stressed that participants should not seek to control the video once they hit the play button. Once all the participants had finished watching the video, the researcher distributed test papers. The participants were told that there was no time limit for completing the test, and they could answer the questions in any order. Finally, the consent form, the questionnaire, the survey, and the test paper completed by each participant were collected and stapled on the upper left corner, completely covering the randomly assigned number.

Scoring

The questions were scored by the researcher and an ESL instructor, both of whom were L1 Chinese speakers. Based on a rubric consisting of 10 acceptable English or L1 idea units in the retention test, an average score for each participant was calculated. Each idea unit was worth 0.5 points and a perfect score was thus 5. For the vocabulary test, there were again 10 idea units and a perfect score of 5. The exact wording was not required, and grammatical mistakes were not considered. Some participants answered the questions in a combination of Chinese and English; no participants used L1 Korean to answer the questions. Inter-rater reliability was high, $r = .992$, $p<.01$ for retention and $r = .987$, $p<.01$ for vocabulary. Minor inconsistencies between the two scorers were resolved through discussion.

Data Analysis

A one-way ANOVA was used to check whether the three groups were comparable. Demographic information including age, first language, years of English learning, months in the US, TOEFL overall scores, and TOEFL reading scores were compared across the three groups. ANCOVA was used to test the effect of three input modes on learning outcomes (retention and vocabulary) controlling for prior knowledge. Before performing the ANCOVA, the assumption of homogeneity was tested and satisfied based on Levene's F-test, $F\ (2, 45) = 2.087$; $p = 0.136$. Normality was tested and met for the comprehension scores based on the Kolmogorov-Smirnov test ($KS_{\text{graphics+audio}}=.185$, $df=17$, $p=.126$; $KS_{\text{graphics+text}}=.164$, $df=15$, $p=.200$; $KS_{\text{graphics+audio+text}}=.188$, $df=13$, $p=.200$). However, normality for the vocabulary scores were violated ($KS_{\text{graphics+audio}}=.323$, $df=17$, $p=.000$; $KS_{\text{graphics+text}}=.283$, $df=15$, $p=.002$; $KS_{\text{graphics+audio+text}}=.307$, $df=13$, $p=.002$). Therefore, Quade's (1967) rank ANCOVA was conducted to examine statistically significant differences among the three groups on the vocabulary scores, controlling for prior knowledge.

Results

Comparability of Groups. ANOVA test examined whether the graphics + audio (N = 17), graphics + text (N = 16), and graphics + audio + text (N = 15) groups differed on any of the characteristics addressed in the questionnaire and the survey. The descriptive statistics calculated are presented in Table 5. ANOVA (with p <. 05) indicated the three groups were not significantly different with regard to age, first language, years of English learning, months in the US, TOEFL overall scores, TOEFL reading scores, TOEFL listening scores, and prior knowledge.

Table 5

Descriptive Statistics

Items	Groups	N	Mean	*SD*	*SE*
Age (years)	graphics+audio	17	19.53	1.772	.430
	graphics+text	16	20.06	2.380	.595
	graphics+audio+text	15	19.53	1.060	.274
	Total	48	19.71	1.810	.261
Years of English learning at school	graphics+audio	15	10.00	3.273	.845
	graphics+text	15	9.67	3.697	.955
	graphics+audio+text	12	10.25	2.221	.641
	Total	42	9.95	3.115	.481
TOEFL overall scores	graphics+audio	16	91.38	6.510	1.628
	graphics+text	11	90.36	6.652	2.006
	graphics+audio+text	14	90.36	6.675	1.784
	Total	41	90.76	6.457	1.008
TOEFL Reading score	graphics+audio	15	24.93	3.195	.825
	graphics+text	11	25.00	2.898	.874
	graphics+audio+text	14	24.71	2.585	.691
	Total	40	24.88	2.839	.449
TOEFL Listening score	graphics+audio	14	24.43	3.155	.843
	graphics+text	11	22.36	2.942	.887
	graphics+audio+text	14	23.43	2.901	.775
	Total	39	23.49	3.042	.487
Prior knowledge test score	graphics+audio	17	3.71	2.664	.646
	graphics+text	16	3.94	2.816	.704
	graphics+audio+text	15	4.73	2.520	.651
	Total	48	4.10	2.652	.383
Retention test score	graphics+audio	17	3.059	1.1164	.2708
	graphics+text	16	3.031	1.2970	.3243
	graphics+audio+text	15	3.033	1.5864	.4096
	Total	48	3.042	1.3080	.1888
Vocabulary test score	graphics+audio	17	.441	.3906	.0947
	graphics+text	15	.400	.3381	.0873
	graphics+audio+text	13	.962	.9233	.2561
	Total	45	.578	.6212	.0926

Learning Results. A one-way ANCOVA indicated that there were no statistically significant differences between the retention scores of the graphics + audio (M = 3.06, SD = 1.12), graphics + text (M = 3.03, SD = 1.30), and graphics + audio + text (M = 3.03, SD = 1.59) groups, $F(2, 45) = .019$, $p = .981$, *partial* η^2=.001. The rank ANCOVA also revealed no significant difference for the vocabulary rank scores, $F(2, 42) = .174$, $p = .841$. Therefore, the three null hypotheses were not rejected. The modality principle and the redundancy principle thus did not appear to have applied to the learning of ESL students. (Liu et al., 2018, pp. 194-197)

Study 2

Study 1 addressed the gap in the literature that no empirical studies directly explored the applicability of the modality and redundancy principles for ESL students. It revealed that both the modality and redundancy principles did not apply for ESL students. However, I was not confident about the results, because Study 1 was limited by the following validity threats.

First, Study 1 only focused on intermediate-level (predominantly L1 Chinese ESL students) and ESL students speaking other L1s at different English-proficiency levels were not included, which is an external validity issue as identified by Mohsen (2016). The number of participants in each input mode was also small, which can be a threat to statistical validity.

Second, the posttest only consisted of a retention test and a vocabulary test. Lack of other types of test to assess learning can be a threat to construct validity. A knowledge transfer test that assesses how well learners use the learned knowledge in new contexts (Mayer et al., 2014) should have been included in the posttest of Study 1.

Third, the assessment of prior knowledge about the target content knowledge might not have been accurate. It was content knowledge rather than language skills that the prior

knowledge survey was designed to assess. However, ESL participants' insufficiency in English language could prevent them from comprehending the survey questions, which creates a threat to construct validity. Therefore, the technical terms in the survey should have been provided with L1 translations.

Fourth, categorical variables of gender, year of study, and native language were not analyzed for their compatibility across the three groups, which is a threat to internal validity.

Last, there were some issues with the implementation of the experiment, which created validity threats. For example, some participants had difficulty locating the multimedia lesson in their assigned mode.

To yield more rigorous findings to the research questions, Study 2 replicated and extended Study 1 by addressing its methodological limitations and implementational issues, which was elaborated in the following sections.

Design

The design of Study 2 was the same as that of Study 1. However, Table 3 utilized in Study 1 did not accurately illustrate the design, because pre-tests in Table 3 were not conducted. Therefore, Table 6 was developed to demonstrate the design of Study 2.

Table 6

Comparisons of Input Modes to Answer the Research Questions

Input Mode	Question 1	Question 2	Question 3
Graphics + Text	✓	✓	✓
Graphics + Audio	✓	✓	
Graphics + Audio + Text	✓		✓

Note: ✓ indicates the input mode involved.

According to Table 6, the following explains how each question would be answered:

1. To answer the first question (i.e. Which of the three input modes is the most facilitative for ESL students' learning?), posttest results from the graphics + audio, graphics + text, and graphics + audio + text groups were compared with each other.
2. To answer the second question (i.e. Does the modality principle apply to ESL students' multimedia learning?), posttest results for the graphics + audio group were compared with those for the graphics + text group.
3. To answer the third question (i.e. Does the redundancy principle apply to ESL students' multimedia learning?), posttest results for the graphics + audio group were compared with those for the graphics + audio + text group.

Participants

Unlike in Study 1, a priori power analysis was first conducted to justify the number of participants in Study 2. According to a review by Mayer and Pilegard (2014), a median effect size of $f = 0.38$ was found for the modality principle. In another review, Mayer and Fiorella (2014) reported a median effect size of $f = 0.43$ for the redundancy principle. Since Study 2 explored the applicability of the modality and redundancy principles for ESL students, an estimated effect size $f = 0.405$ was used in a priori power analysis on G*power 3.1. Based on the power analysis, to achieve a generally acceptable power of 0.8, the estimated sample size was 63 with an estimated 21 participants in each of the three conditions.

In Study 2, all participants in the study were from a research university in the northeastern part of the US. To address the external validity threat in Study 1 where ESL students were recruited from the same intermediate-level ESL classes, Study 2 explored the research questions with a more representative population that consisted of ESL students at uncontrolled

proficiency levels. Therefore, unlike Study 1 that recruited its participants from ESL classes of the same immediate-level, Study 2 utilized the following two methods to recruit more diverse ESL participants:

1. An advertisement about the study was written in English and posted on the notice boards of classroom buildings and libraries of the university, as well as in the group chat room for international students on WECHAT(i.e. a popular social media app). Two students recruited from WECHAT did not take the posttest, and therefore were not counted. Altogether 20 ESL students were recruited from the advertisement: 19 from notice boards and one from WECHAT. The students were promised $8.00 to participate in the experiment, payable after they completed the experiment.
2. A communications professor who is a friend of the researcher let him solicit ESL participants in her classes. Forty ESL students were recruited from the professor's classes. Those ESL students who finished the experiment were given extra credit as reward.

The following lists basic information about the recruited participants:

1. Compared with the 48 in Study 1, altogether 60 ESL participants were recruited during a period of eight months.
2. Compared with participants' average age of 19.71 in Study 1, the average age was 22.80 (*SD*=3.16).
3. Compared with 24 males and 24 females in Study 1, there turned out to be 16 males and 44 females recruited.
4. As in Study 1, the majority of the participants turned out to be native Chinese speaking, 56 of them (93.3%) identified Mandarin or Chinese as their first language;

three identified Korean as their first language; one spoke Hindi as the first language, although English was the instructional language during his schooling in India. The four participants whose L1s were not Chinese were recruited from the professor's classes.

5. The 60 ESL participants were evenly and randomly assigned to the three input mode conditions. Each Korean student was assigned to one condition. Twenty participants were in the graphics + audio group, 20 were in the graphics + text group, and 20 were in the graphics + audio + text group.

Instructional Material

The instructional material of both Study 1 and Study 2 was based on a YouTube animation (Moulton, 2010). This animation was a five-minute colored video clip teaching the formation of lightning. It began with a brief introduction of the five steps of lightning, and then a detailed description of each step. A geography professor was consulted to ensure the accuracy of the content knowledge taught in the animation. It was selected for two reasons.

First, it generally satisfied the requirements brought about by the research questions. More specifically, it explicitly taught the mechanism about the formation of lightning, a typical example of content knowledge, and it contained some explicitly taught technical terms.

Second, with the rapid development of multimedia technology, a color video rather than a black-and-white slideshow used by Moreno and Mayer (2002) is now more likely to be used for instruction. Last, it was generally an effective multimedia lesson, since it followed the pre-training principle, the spatial contiguity principle, the signaling principle, the multimedia principle, the image principle (Mayer, 2009), as well as the modified redundancy principle (Mayer & Johnson, 2008). The principles ensured that the animation promoted content

knowledge learning.

However, the original animation needed to be modified to be able to examine the research questions. It was modified as follows to become the multimedia lesson.

First, to compare the input modes in the modality and redundancy principles that include on-screen text, transcription of the auditory input was added to the animation. Text in black Times New Roman font was at the bottom of the multimedia lesson, which is conventionally the way text or caption is presented. The transcription of auditory input is provided in Appendix A.

Second, since narration and graphics of the original animation did not completely synchronize, the temporal principle (Mayer, 2009) was not followed; since some graphics in the original animation was unrelated to the content, the coherence principle (Mayer, 2009) was not followed. The dyssynchronization and the unrelated graphics did not promote learning, and therefore the original animation was modified accordingly:

1. Narration in the animation was adjusted to be completely synchronous with its corresponding graphics;
2. Irrelevant graphics in the animation, such as a large eye that covers the whole screen, were deleted.

The instructional material of Study 2 was based on that of Study 1. However, certain modifications were made based on participants' feedback and researcher's observation in Study 1. The feedback was received via informal conversations with about 10 participants immediately after the implementation of Study 1. Additionally, the rationale for controlling for the instructional material that was not elaborated in Study 1 was further explained.

Modifications of Study 1. To better examine the research questions in Study 2, the following modifications were made to the instructional material of Study 1.

First, unlike in Study 1, a five-second red silent screen consisting of a question: *What is lightning?* and a numeral of one, two or three (each representing one input mode) preceded the introduction. The silent screen was illustrated in Figure 4. The added silent screen aimed at better helping participants locate their assigned multimedia lessons on YouTube.

Second, the speed of the multimedia lesson of Study 1 was adjusted to 90 percent of the original speed in Study 2. Some participants in Study 1 reported during informal conversations with the researcher that the narration was faster than the normal speech speed, which was confirmed by one native English-speaking and one ESL doctoral students majoring in Education. They both agreed that 90 percent of the original play speed was more suitable for instruction.

Figure 4. The silent screen on the multimedia lesson in input mode 2.

Rationale of controlling for the instructional material. As explained in Chapter 2, not any random video clip is appropriate for instructional purposes and suitable to be the multimedia material of Study 2. The reviewed empirical studies indicate, "very few studies have provided a way by which material could be assessed for suitability for in-class or self-study activities." (Liversidge, 2000, p. 22-23), and they were "not done with material from an existing academic course, and as such, the lesson itself may have not been as conducive to learning as a well-refined lecture may be" (Mayer et al., 2014, p. 659). Only multimedia lessons controlled to be

conducive to learning should be used for instruction in the first place, which became the premise for the discussion about different input modes for ESL students learning via multimedia lessons.

In Study 2, Mayer's multimedia learning principles (2009) were used to control the multimedia lesson of Study 2 for the following three reasons:

1. They ensured the multimedia lesson to be conducive to learning.
2. They ensured that the multimedia lesson was representative of different multimedia lessons, because a single case scenario without following certain principles is insufficient to represent the whole construct (Bellini & Rumrill, 2009).
3. They met the conditions for the modality and redundancy principles. As explained in Chapter 2, certain multimedia learning principles can influence the applicability of the modality and redundancy principles. To examine the applicability of the two principles for ESL students, their conditions should be met; otherwise, conditions
4. could be a variable threatening internal validity.

In the multimedia lesson of Study 1, three multimedia learning principles of the multimedia principles, the personalization (Mayer, 2009), and the modified redundancy principles (Mayer & Johnson, 2008) were all followed, but they were not explicitly explained. In the multimedia lesson of Study 2, nine learning principles including the multimedia principle, the coherence principle, the pre-training principle, the spatial contiguity principle, the temporal contiguity principle, the signaling principle, the voice principle, the image principle, and the personalization principle (Mayer, 2009), as well as the modified redundancy principle (Mayer & Johnson, 2008) were controlled to be followed to investigate the modality and redundancy principles for ESL students. The native English-speaking doctoral student and the ESL doctoral students who confirmed the play speed of the multimedia lesson also verified the application of

the principles. The 10 principles were illustrated in Table 7.

Of the 12 multimedia learning principles (Mayer, 2009), the segmenting principle that people learn better when they can control the pace of the multimedia lesson was found to be able to counteract the modality and redundancy principles (Cheon et al., 2014). Not following the segmenting principle is the condition for the original modality and redundancy principle to apply. Therefore, to investigate the applicability of the modality and redundancy principle for ESL students, the segmenting principle was not followed in the multimedia lesson.

Table 7

Ten Multimedia Principles Observed in the Design of the Multimedia Lesson (adapted from Mayer, 2009)

Name of the Principle	Explanation of the Principle
The coherence principle	Content irrelevant to learning is removed from the animation.
The spatial contiguity principle	On-screen keywords are placed near the corresponding pictures.
The temporal contiguity principle	Graphics and corresponding narrations are presented simultaneously.
The pre-training principle	Keywords and key steps are explained briefly before receiving treatments.
The modified redundancy principle	Keywords and key concepts are placed next to the corresponding portion of the diagram.
The voice principle	People learn better when the narration is spoken in a standard-accented human voice than a machine voice or foreign-accented human voice.
The image principle	Presence of the image of the speaker does not promote learning.
The multimedia principle	Graphics and words are better than words alone for learning.
The signaling principle	People learn better when cues are added to highlight the organization of target content knowledge.
The personalization principle	People learn better when words are delivered in a conversational style.

Instruments

As in Study 1, the instruments in Study 2 (Appendix B) consisted of a demographic

questionnaire, a prior knowledge survey, and a posttest. Since this study attempted to extend Mayer's (2009) original modality and redundancy principles by addressing their applicability to ESL students' learning, the questionnaire, the prior knowledge survey, and the posttest were adopted from studies (e.g. Mayer & Moreno, 1998; Mayer et al., 2014; Moreno & Mayer, 1999, 2002) examining similar questions, and then modified for this study.

The demographic questionnaire, the prior knowledge survey, and the posttest were individually described, as Study 1 did not sufficiently explain them. To better assess learning for Study 2, certain modifications were made to the instruments of Study 1. Additionally, the rationale for providing L2 translation that was not elaborated in Study 1 was further explained.

The demographic questionnaire. The same demographic questionnaire in Study 1 was used for Study 2. Based on the questionnaire made by Mayer et al. (2014) who examined input modes and ESL students' content learning, the questionnaire (in Appendix B) was designed to elicit participants' basic demographic and English background information.

To gather more comprehensive background information about ESL participants, months in the US and years of English learning at school that could have an influence on English proficiency (Lightbown & Spada, 2013; Macaro, 2010) were added to the questionnaire of Mayer et al. (2014). The questionnaire of Study 2 then included gender, age, native language, year in college, months in the US, years of English learning at school, SAT score, TOEFL reading score, and TOEFL listening score.

The prior knowledge survey. The same prior knowledge survey in Study 1, included in Appendix B, was used for Study 2. It aimed at examining the compatibility of participants in terms of their background knowledge about the content to be learned. It was based on the survey of Mayer and Moreno (1998) and that of Moreno and Mayer (1999; 2002), because the

animation utilized by Mayer and Moreno (1998) and Moreno and Mayer (1999; 2002) taught university students the same content about the formation of lightning as the multimedia lesson of Study 1 and Study 2.

The prior knowledge survey of Study 1 and Study 2 consisted of nine question. In the first eight questions, participants were asked if they understood some specific concepts about meteorology. There were eight statements and participants were asked to check the one(s) that applied to them. In the ninth question, participants were asked to evaluate their own general knowledge about meteorology. They rated their knowledge on a scale of 1-5, 1 being very little and 5 being very much. However, different from the survey used by Mayer and Moreno (1998), Moreno and Mayer (1999; 2002), the prior knowledge survey of Study 1 was provided with L1 translations of the directions and keywords, such as *meteorology*, to avoid participants' lack of English proficiency influencing their understanding.

Additionally, based on participants' feedback about the survey of Study 1, L1 translations of more technical terms were provided in the survey of this study. Two participants of Study 1 reported after the experiment that they did not recognize the words: *cumulous* and *nimbus*, but they knew the meanings of the two words in their native language. It was the meaning (rather than the form) of the technical term that the survey was designed to assess. Therefore, more L1 translations of technical terms were added to ensure full comprehension.

The posttest. The posttest of Study 2 (in Appendix B) consisted of a knowledge retention test, a knowledge transfer test, and a vocabulary test. Mayer and Moreno (1998), Moreno and Mayer (1999; 2002), and Mayer et al. (2014) explored the original modality and redundancy principles, and they suggested that learning assessments should address both knowledge retention and transfer. Therefore, to examine the applicability of the modality and redundancy

principles for ESL students, the posttest of Study 2 also focused on assessing both knowledge retention and knowledge transfer.

The knowledge retention test and the vocabulary test in Study 2 were based on those in Study 1, and certain modifications were made. A knowledge transfer test absent in Study 1 was developed and added to the posttest of Study 2.

The retention test. The same retention test of Study 1 was used in Study 2. The retention test was to assess how well the new content knowledge was retained after the treatment. Mayer and Moreno (1998), Mayer, Moreno, Boire, and Vagge (1999), Moreno and Mayer (1999; 2000) and Mayer et al. (2014) all examined input modes for content knowledge learning. They required the participants to simply write down what they could remember about the multimedia lesson in English. However, the retention test (in Appendix B) of Study 1 and Study 2 required participants to recall, explain the five steps of lightning, as well as illustrate the five steps on five pictures (Figure 5).

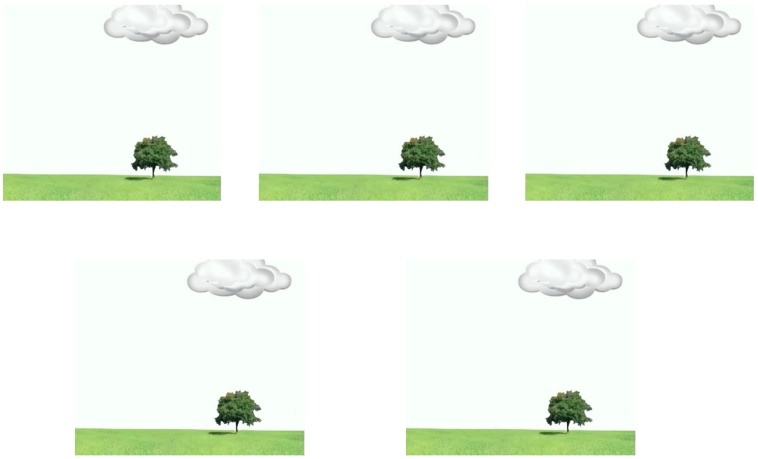

Figure 5. The illustration portion of the retention test.

In Study 1 and Study 2, asking participants to explain in steps that were explicitly taught in the multimedia lesson aimed at promoting production. The illustration portion complemented the explanation portion. In it, participants illustrated the five steps of lightning on the same

picture consisting of grassland, a tree and a cloud. The reasons of adding an illustration portion were two-fold. First, the illustration portion did not require language demand, which encouraged ESL students' production. Second, adding more questions to the retention test could increase its sensitivity and accuracy (Mayer, 2011).

However, based on the feedback and test performances of some participants in Study 1, the directions were modified by including an explanation that all the answers should be based on the information in the multimedia lesson rather than prior knowledge or guessing. To address construct validity, the geography professor was consulted to ensure the accuracy of the questions.

The vocabulary test. A vocabulary test (in Appendix B) was designed to assess participants' retention of the meanings of some explicitly taught technical terms. Since different formats of vocabulary tests tap different aspects of vocabulary knowledge (Kieffer & Lesaux, 2012), the vocabulary test in Study 1 and Study 2 focused on highly context-specific terminology and assessed ESL students' identification of form and retention of meaning. In both Study 1 and Study 2, participants were allowed to explain the meaning of the technical term in their L1s or English, or both.

Compared with the vocabulary test in Study 1, there was one more question item (i.e. *neutralization*) on the vocabulary test in Study 2. Because five technical terms were explicitly taught in the multimedia lesson, there were five vocabulary questions (i.e. *updraft, downdraft, step-leader, streamer, and neutralization*) in Study 2. Additionally, based on the feedback from some participants, as well as their test performances in Study 1, the directions were modified by adding an explanation that definitions of the terms should be solely based on the multimedia lesson. To address construct validity, the geography professor was consulted to ensure the accuracy of the questions.

The transfer test. To more comprehensively assess learning, a knowledge transfer test (in Appendix B) that was absent in Study 1 was added to the posttest of Study 2. As Mayer and Moreno (1998), Moreno and Mayer (1999; 2002), and Mayer et al. (2014) suggested, content knowledge learning should address both knowledge retention and transfer, and transfer test questions should examine how well participants applied the learned content knowledge to new contexts. Three transfer questions were developed based on the three key concepts explicitly taught in the multimedia lesson. To address construct validity, the transfer test was developed in consultation with the geography professor, ensuring the accuracy of the questions. Minor disagreements on wording were resolved through discussion. There three questions are:

1. *Before lightning, is an airplane flying on top of the clouds positively or negatively charged? And why?*
2. *When you are on an open area and see lightning in the near distance, is it safe to hide in a ditch? If it is not, what should you do?*
3. *Why do we sometimes see heavy clouds in the sky but no lightning?*

For each transfer question, participants read about a situation, then made a judgment based on the information learned from the multimedia lesson, and last explained the rationale for the judgment. ESL participants' lack of English proficiency prevents them from understanding test questions or producing answers (Rhodes et al., 2005; Spinelli, 2008). Therefore, both test questions and directions were also translated into L1s, and participants were allowed to use L1s to answer questions.

Summary of modifications. Based on the discussed limitations, the feedback from participants, and researchers' observation during the implementation of Study 1, modifications were made to the instruments of Study 1 to better investigate the research questions. They were:

1. More L1 translations of keywords were provided in the prior knowledge survey.
2. An explanation was added to the directions of the retention and vocabulary tests.
3. The vocabulary test consisted of more vocabulary items.
4. A knowledge transfer test was added to the posttest.

The retention test, the transfer test, and the vocabulary test each was on a separate page. The three pages were stapled in a random order, the purpose of which was to eliminate the possible external validity threat brought about by the order of administering the three tests.

Rationale for providing L2 translation. In both Study 1 and Study 2, it was content knowledge rather than language that the instruments were designed to assess. Participants' lack of English language proficiency could create barriers to understanding test questions (Rhodes, Ohoa, & Ortiz, 2005; Spinelli, 2008). The following two measures enabled participants to fully understand all question items on the instruments.

First, unlike in Study 1, an ESL instructor was consulted in Study 2 to provide feedback on the language use of the test items. The feedback ensured that the English language use was natural and simple in sentence structures and word choice, and free of social-cultural references that might prevent ESL students from accurately comprehending questions.

Second, unlike Mayer et al. (2014) who neither provided their ESL participants with L1 translations of test questions, nor allowed them to use L1 for answering questions in their posttest, the researcher provided L1 translations of directions and questions. Additionally, ESL participants in this study were allowed to use L1s to answer questions, the purpose of which was to ensure that ESL students' lack of English language skills would not influence their production.

Google translation was utilized to translate technical terms, question items, as well as the directions into Chinese and Korean beforehand. Therefore, the instruments were written in

English and native language translations. Only Chinese and Korean versions of the instruments were available, because ESL students on campus were overwhelmingly native Chinese or native Korean speaking, which was consistent with the demographics of participants in Study 1. Two doctoral students whose L1s were Chinese and Korean were consulted to ensure the accuracy of the translation. Participants who spoke other L1s were allowed to use Google translation during the experiment.

Procedure

Based on researchers' observation and participants' feedback, there were some issues during the implementation of the experiment in Study 1. The issues were addressed in Study 2 .

Issues with the procedure of Study 1. First, some participants reported during the informal conversations with the researcher that they did not expect that they would not hear narration from their headphones, and they thought that their headphones were not working properly, so they spent time adjusting their headphones after the multimedia lesson began, which reduced some of their learning time. Since one input modes was graphics + text, participants assigned to this input mode condition would not hear auditory input.

Secondly, based on researcher's observation during the experiment, some participants did not pay enough attention to the announcement that they would watch the lesson only once and there was a time limit for learning. They thought that they would be able to watch the lesson more than once, so they did not give it a full-attention.

Thirdly, some participants took notes during watching, and some of them paused the lesson while taking notes, although they were clearly instructed not to control the pacing of the lesson.

The procedure of Study 2. The experiment of Study 2 was conducted in six sessions over a period of 10 months. The six sessions included two sessions of participants recruited from the advertisement and four sessions of participants recruited from the professor's classes. The number of each session varied. More specifically, there were 11, 9, 12, 15, and 13 participants in each session respectively. Twenty participants were randomly assigned to each of the three input conditions. The following procedure was the same for each session. In a classroom, participants were informed of the purpose of the study and given consent forms to sign. The consent form explained that participation was voluntary and the data to be collected were kept private and anonymous. All participants signed the content forms. After the study was introduced to the ESL participants, participants were asked if they heard about this study before to eliminate the possible contamination effect brought about by interacting with previous participants. None of the participants reported that they did.

Then the researcher distributed the questionnaire and the survey. The participants were given the versions with their corresponding L1 translations. Since no Hindi version of the instruments was available, the native Hindi-speaking participant was allowed to use Google translation during the experiment. However, he reported that he understood everything on the instruments, and therefore he did not use translation support.

After all participants completed the questionnaires and the surveys, they were directed to the bottom of the survey. Each participant was randomly assigned a keyword on the bottom of the survey. Each keyword (i.e. yinanliu1, yinanliu2, and yinanliu3) signified one type of input modes. Then the participants were instructed to turn on their laptops that they were requested to bring, open YouTube, and use the keywords to locate their assigned multimedia lessons. The researcher instructed that all participants stopped there once they found their multimedia lessons.

When all participants had located their assigned lessons and stopped, the researcher announced to all participants that:

1. They would watch a video lesson with one of the three input modes: graphics + audio, graphics + text, and graphics + audio + text. Some students might not be hearing narration in their headphones, because they were assigned to watch the lesson with the mode of graphics + text.
2. They would watch the video lesson only once and need to give it full attention for five minutes.
3. Although they were allowed to take notes during watching, they were not encouraged to do so. The video lesson was information-loaded and taking notes might lead to distraction, and therefore a loss of information.

These announcements were based on researcher's observation in Study 1, and aimed at eliminating confusion among participants. When all participants were ready, the researcher instructed them to put on their headphones and start playing the video.

After five minutes, the researcher instructed the participants to close their laptops, and then distributed the posttest. Five minutes were enough for participants to finish watching the tutorial based on the observation during Study 1. The participants were informed that the posttest would be collected with the questionnaire and the survey after 10 minutes. Ten minutes were more than enough based on the observation of Study 1. After 10 minutes, the questionnaire, the survey, and the posttest paper of each participant were collected, and then the participants were dismissed. The experiment lasted for approximately 25 minutes for each session.

Scoring and Data Analysis

Scoring in Study 2 was based on that in Study 1. However, for Study 2, certain modifications were made to the data analysis of Study 1 to address the limitation that categorical variables were not analyzed. Below are more detailed descriptions of scoring and date analysis in Study 2.

Scoring. In Study 2, the researcher created a data file for the project on SPSS (Version 23). Each participant was assigned a number. There were 15 variables: assigned number, age, gender, native language, year of study, months in the US, years of English learning at school, TOEFL Reading score, TOEFL listening score, prior knowledge survey result, input modes, self-evaluation survey, retention test score, vocabulary test score, and transfer test score. For each participant, all responses to the question items on the demographic questionnaire were recorded; for the first eight questions on the prior knowledge survey, the sum of all the checked items was inputted, so was the selected number to the self-evaluation question.

The researcher and a native Chinese-speaking doctoral student each rated all the tests. The other rater was trained first, which ensured that the two raters had a consensus about the rubric. The posttest rubric consisted of 20 acceptable idea units for the retention test, 10 for the vocabulary test, and 6 for the transfer test. Each idea unit was worth one point, and there was no half point. Then the perfect score was 20 for the retention test, 10 for the vocabulary test, and six for the transfer test. For each participant, the retention test, the transfer test and the vocabulary test were scored based on the total number of acceptable idea units. Exact wording was not required. Grammatical mistakes and errors were not taken into account. Some participants answered the questions in both Chinese and English; no participants used L1 Korean to answer the questions. Inter-rater reliability was high, $r = 0.91$, $p < .01$ for retention, $r = 0.92$, $p < .01$ for

vocabulary, and $r = 0.89$, $p < .01$ for transfer. Minor inconsistencies in scores between the two raters were resolved through discussion.

Data analysis. In study 1, categorical variables of gender, year of study, and native language were not analyzed for comparability across the three groups. To address this limitation, in Study 2, Pearson's chi-square tests were used for comparability analysis of the three categorical variables in the questionnaire.

In Study 2, all statistical tests were performed using SPSS (Version 23). An alpha level of .05 was used for all subsequent analyses. Before comparing learning outcomes, the researcher utilized a one-way ANOVA test to determine if participants in the three conditions were comparable in terms of some basic characteristics including age, SAT scores, TOEFL scores and prior knowledge. Pearson's chi-square tests examined the comparability of categorical variables including gender, year of study, native language, and the one self-evaluation question (i.e. rate general knowledge about meteorology on a scale of 1 to 5) in the prior knowledge survey across the three groups.

Then, the researcher utilized ANOVA tests to compare retention test scores, transfer test scores and vocabulary test scores in the three conditions. When participants in the three groups were not comparable for certain variables in demographics and prior knowledge, ANCOVA tests would be used instead. Before performing ANOVA or ANCOVA tests, the assumptions of ANOVA or ANCOVA (i.e. normality and homogeneity) were examined by the Shapiro-Wilk test and the Levene's test. When normality was violated, nonparametric Kruskal-Wallis tests were performed to compare test scores in the three conditions.

Chapter 4 elaborates the results of Study 2 generated by the above data analysis method.

CHAPTER 4

RESULTS OF STUDY 2

In this chapter, the three treatment groups in Study 2 were compared to examine comparability in terms of demographics and prior knowledge. Then, test results of the three groups were compared to answer the three research questions. The research questions are:

1. Which of the three input modes is the most facilitative for ESL students' learning?
2. Does the modality principle apply to ESL students' multimedia learning?
3. Does the redundancy principle apply to ESL students' multimedia learning?

Group Comparability

Before test results of the three treatment groups were compared to answer the research questions, it was necessary to examine if three groups were comparable in terms of their demographics and prior knowledge. For statistical analysis, there were two types of variables: categorical variables (i.e. gender, years of study, native language, and self-evaluation of prior knowledge) and non-categorical variables (i.e. age, SAT scores, TOEFL reading scores, TOEFL listening scores, months in the US, years of English learning, and prior knowledge survey scores).

Comparability of Categorical Variables

For categorical variables including gender, year of study, native language, and self-evaluation in the prior knowledge survey, Pearson's chi-square tests were performed to examine the comparability of the three groups.

Test results indicated that the three groups did not differ significantly on gender, χ^2 (2, N = 60) = 2.73, p = .26, year of study, χ^2 (8, N = 56) = 5.53, p = .70, native language, χ^2 (4, N = 60) = 2.04, p = .73, or self-evaluation score in prior knowledge survey, χ^2 (6, N =60) = 11.03, p = .09, because each p value for gender (p = .26), year of study (p = .70), native language (p = .73), and self-evaluation of prior knowledge (p = .09) was over the alpha of .05.

Comparability of Non-Categorical Variables

Table 8

Descriptive Statistics of Non-Categorical Variables in Study 2

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
age (years)	video+audio	20	22.85	2.254	.504	21.79	23.91	18	27
	video+text	20	22.40	1.984	.444	21.47	23.33	18	27
	video+audio+text	20	23.15	4.660	1.042	20.97	25.33	18	35
	Total	60	22.80	3.161	.408	21.98	23.62	18	35
SAT	video+audio	3	1900.00	173.205	100.000	1469.73	2330.27	1800	2100
	video+text	2	1785.00	120.208	85.000	704.97	2865.03	1700	1870
	video+audio+text	4	1957.50	187.683	93.842	1658.85	2256.15	1750	2200
	Total	9	1900.00	165.756	55.252	1772.59	2027.41	1700	2200
months in the US	video+audio	16	17.44	17.694	4.423	8.01	26.87	4	58
	video+text	17	15.71	18.764	4.551	6.06	25.35	1	60
	video+audio+text	12	28.08	19.902	5.745	15.44	40.73	4	60
	Total	45	19.62	18.999	2.832	13.91	25.33	1	60
years of English learning at school	video+audio	11	13.82	5.862	1.767	9.88	17.76	5	24
	video+text	16	13.25	4.266	1.067	10.98	15.52	5	20
	video+audio+text	9	12.67	4.153	1.384	9.47	15.86	6	20
	Total	36	13.28	4.664	.777	11.70	14.86	5	24
TOEFL Reading score	video+audio	9	26.78	1.986	.662	25.25	28.30	24	30
	video+text	12	26.00	2.629	.759	24.33	27.67	23	30
	video+audio+text	7	26.86	3.288	1.243	23.82	29.90	21	30
	Total	28	26.46	2.560	.484	25.47	27.46	21	30
TOEFL Listening score	video+audio	9	25.00	2.646	.882	22.97	27.03	21	30
	video+text	13	25.00	3.629	1.006	22.81	27.19	17	30
	video+audio+text	7	24.43	3.599	1.360	21.10	27.76	18	29
	Total	29	24.86	3.237	.601	23.63	26.09	17	30
prior knowledge survey score	video+audio	20	2.80	1.704	.381	2.00	3.60	0	6
	video+text	20	2.30	2.179	.487	1.28	3.32	0	6
	video+audio+text	20	2.45	2.235	.500	1.40	3.50	0	7
	Total	60	2.52	2.029	.262	1.99	3.04	0	7

The descriptive statistics of non-categorical variables were presented in Table 8. Before ANOVA tests were performed to examine the comparability of the three groups in terms of age, SAT scores, TOEFL reading scores, TOEFL listening scores, months in the US, years of English learning, and prior knowledge survey scores, the assumptions of ANOVA (i.e. normality and

homogeneity) were examined. More specifically, Shapiro-Wilks tests were performed to examine the assumption of normality, and Levene' tests were performed to examine the assumption of homogeneity. For those variables that did not meet the assumptions, nonparametric Kruskal-Wallis tests were performed to examine comparability.

Normality tests. Normality was satisfied for TOEFL reading scores, $SW_{\text{graphics+audio}} = 0.96$, $df = 9$, $p = .85$; $SW_{\text{graphics+text}} = 0.89$, $df = 12$, $p = .11$; $SW_{\text{graphics+audio+text}} = 0.88$, $df = 7$, $p = .24$, TOEFL listening scores, $SW_{\text{graphics+audio}} = 0.93$, $df = 9$, $p = .47$; $SW_{\text{graphics+text}} = 0.94$, $df = 13$, $p = .48$; $SW_{\text{graphics+audio+text}} = 0.96$, $df = 7$, $p = .85$, and years of English learning, $SW_{\text{graphics+audio}} = 0.90$, $df = 11$, $p = .17$; $SW_{\text{graphics+text}} = 0.92$, $df = 16$, $p = .20$; $SW_{\text{graphics+audio+text}} = 0.97$, $df = 9$, $p = .88$, since all p values were above .05. However, the assumption of normality was not satisfied for age, SAT scores, months in the US, and prior knowledge survey, which means that ANOVA tests were not appropriate for the subsequent analyses.

Homogeneity tests. Levene' tests were performed to examine the assumption of homogeneity for the three variables that satisfied normality: TOEFL reading scores, TOEFL listening scores, and years of English learning. Levene' test results indicated that the assumption of homogeneity for TOEFL reading scores ($p = .43$), TOEFL listening scores ($p = .93$), and years of English learning ($p = .87$) was all met, since the p values were all over .05. Therefore, ANOVA tests were appropriate for the subsequent analyses regarding TOEFL reading scores, TOEFL listening scores, and years of English learning.

ANOVA tests. Since both normality and homogeneity, the two assumptions of ANOVA, were satisfied for TOEFL reading scores, TOEFL listening scores, and years of English learning, ANOVA tests were performed to examine the comparability of the three groups in terms of the

three variables. ANOVA tests in Table 9 revealed that participants in the three groups did not differ significantly in terms of TOEFL reading scores, F (2, 25) = .33, p = .72, TOEFL listening scores, F (2, 26) = .08, p = .93, and years of English learning, F (2, 33) = .14, p = .87, since the p values were all over .05.

Table 9

ANOVA Tests for TOEFL Reading and Listening Scores, and Years of English Learning

		Sum of Squares	df	Mean Square	F	Sig.
TOEFL Reading scores	Between Groups	4.552	2	2.276	.330	.722
	Within Groups	172.413	25	6.897		
	Total	176.964	27			
TOEFL Listening scores	Between Groups	1.734	2	.867	.077	.926
	Within Groups	291.714	26	11.220		
	Total	293.448	28			
Years of English learning at school	Between Groups	6.586	2	3.293	.144	.866
	Within Groups	754.636	33	22.868		
	Total	761.222	35			

Nonparametric tests. For age, SAT scores, months in the US, and prior knowledge survey scores that did not satisfy the assumptions of normality and homogeneity, ANOVA tests were not appropriate. Instead, nonparametric tests were performed. According to Kruskal-Wallis tests, age (p = .60), SAT scores (p = .45), months in the US (p = .18), and prior knowledge survey scores (p = .53) were not statistically different across the three groups, since all the p values were over .05.

Based on the above analysis, participants in the three groups were not statistically significant in terms of all the variables in the demographics questionnaire and the prior knowledge survey. Therefore, the participants in the three groups were comparable.

Learning Results

The descriptive statistics of outcome variables were presented in Table 10. As illustrated in Appendix B, there were 10 questions in the retention test, five in the vocabulary test, and three in the transfer test. Each question included two acceptable idea units, and each question was then

worth two points.

Table 10

Descriptive Statistics of Outcome Variables

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
retention test score	video+audio	20	13.50	4.594	1.027	11.35	15.65	4	20
	video+text	20	11.15	4.804	1.074	8.90	13.40	3	19
	video+audio+text	20	11.95	4.571	1.022	9.81	14.09	4	19
	Total	60	12.20	4.682	.605	10.99	13.41	3	20
vocabulary test score	video+audio	20	3.85	3.514	.786	2.21	5.49	0	9
	video+text	20	2.50	2.090	.467	1.52	3.48	0	7
	video+audio+text	20	3.15	3.281	.734	1.61	4.69	0	9
	Total	60	3.17	3.026	.391	2.38	3.95	0	9
transfer test score	video+audio	20	2.15	1.631	.365	1.39	2.91	0	5
	video+text	20	2.50	1.821	.407	1.65	3.35	0	6
	video+audio+text	20	2.00	1.376	.308	1.36	2.64	0	5
	Total	60	2.22	1.606	.207	1.80	2.63	0	6

Reliability of Test Measures

The posttest included the retention test, the vocabulary test, and the transfer test. Cronbach's Alpha was computed to examine the reliability of each test measure. The retention test consisted of 10 items ($\alpha = 0.88$), the vocabulary test consisted of five items ($\alpha = 0.78$), and the transfer test consisted of three items ($\alpha = 0.71$). According to George and Mallery (2009), each test ($\alpha > 0.7$) of Study 2 could be accepted as reliable.

Comparisons of Learning Results

Before ANOVA tests were performed to compare the learning results of retention, vocabulary, and transfer tests across the three groups, the assumptions of ANOVA (i.e. normality and homogeneity) were examined for each test. Specifically, Shapiro-Wilks tests were performed to examine the assumption of normality for retention, vocabulary, and transfer test results; Levene' tests were performed to examine the assumption of homogeneity for retention, vocabulary, and transfer test results. For those that did not meet the assumptions of normality and

homogeneity, ANOVA tests were not appropriate, and therefore nonparametric Kruskal-Wallis tests were performed.

Normality tests. The assumption of normality for retention test scores, vocabulary test scores, and transfer test scores across the three groups were individually examined by Shapiro-Wilk tests. Test results indicated that normality was satisfied for retention test scores, $SW_{\text{graphics+audio}} = 0.93$, $df = 20$, $p = .15$; $SW_{\text{graphics+text}} = 0.96$, $df = 20$, $p = .49$; $SW_{\text{graphics+audio+text}} = 0.95$, $df = 20$, $p = .32$, and transfer test scores, $SW_{\text{graphics+audio}} = 0.92$, $df = 20$, $p = .09$; $SW_{\text{graphics+text}} = 0.93$, $df = 20$, $p = .17$; $SW_{\text{graphics+audio+text}} = 0.92$, $df = 20$, $p = .09$, since the p values were all above .05. However, normality was not satisfied for vocabulary test scores, $SW_{\text{graphics+audio}} = 0.85$, $df = 20$, $p = .01$; $SW_{\text{graphics+text}} = 0.87$, $df = 20$, $p = .01$; $SW_{\text{graphics+audio+text}} = 0.83$, $df = 20$, $p = .01$, since the p values were all below .05.

Homogeneity tests. Levene' tests were performed to examine the assumption of homogeneity for the two variables that satisfied normality: retention test scores and transfer test scores. Based on Levene' tests, the assumption of homogeneity for both retention test scores ($p = .89$) and transfer test scores ($p = .25$) was met, since the p values were both above .05.

Table 11

ANOVA Tests for Retention Test Scores and Transfer Test Scores

		Sum of Squares	df	Mean Square	F	Sig.
retention test scores	Between Groups	57.100	2	28.550	1.316	.276
	Within Groups	1236.500	57	21.693		
	Total	1293.600	59			
transfer test scores	Between Groups	2.633	2	1.317	.502	.608
	Within Groups	149.550	57	2.624		
	Total	152.183	59			

ANOVA tests. For retention test scores and transfer test scores that satisfied the assumptions of normality and homogeneity, ANOVA tests were conducted. Test results in Table

10 revealed that there were no statistically significant differences across the three groups in terms of retention test scores, F (2, 57) =1.32, p = .28, *partial* η^2=.04 and transfer test scores, F (2, 57) = .50, p = .61, *partial* η^2=.02 since the p values were both over .05.

Nonparametric test. According to the nonparametric Kruskal-Wallis test, the vocabulary test scores (p = .74) were not statistically different across the three groups, since the p value was above .05. Therefore, ESL students' retention scores in the three conditions were not statistically different, so were their vocabulary scores, so were their transfer scores.

Summary

Based on the above analyses, in Study 2, the three groups were comparable in terms of demographics and prior knowledge, and there were no statistically significant differences among the three groups in terms of retention test scores, vocabulary test scores, and transfer test scores. Therefore, input modes (i.e. graphics + audio, graphics + test, and graphics + audio + text) did not have an impact on ESL students' learning. None of the null hypotheses were rejected. Specifically,

1. ESL students' learning from the three input modes of graphics + audio, graphics + test, and graphics + audio + text was statistically the same.
2. ESL students' learning from the input mode of graphics + audio was statistically the same as that from the input mode of graphics + text. Therefore, the modality principle (i.e. the mode of graphics + audio is more effective than that of graphics + text) did not apply to ESL students' learning.
3. ESL students' learning from the input mode of graphics + audio was statistically the same as that from the input mode of graphics + audio + text. Therefore, the

redundancy principle (i.e. the mode of graphics + audio is more effective than that of graphics + audio + text) did not apply to ESL students' learning.

CHAPTER 5

DISCUSSION

In this chapter, first, the findings of both Study 1 and Study 2 were summarized, followed by discussions of the findings. Next, the strengths and limitations of Study 2 were explored. Last, a conclusion was drawn and future studies were suggested.

Main Findings

Based on the cognitive theory of multimedia learning (Mayer, 2009), verbatim on-screen text (as in graphics + text, and graphics + audio + text) can impede learning by overloading the visual channel, while auditory input (as in graphics + audio) that complements visual input can facilitate learning. Therefore, graphics + audio is expected to be the most facilitative input mode. However, findings of both Study 1 and Study 2 indicated that it was not the case for ESL students.

In both Study 1 and Study 2, the multimedia lesson was controlled to ensure its conduciveness to learning, which was the premise for the investigation into the three input modes for ESL students' learning. To control the multimedia lesson, the coherence principle, the pre-training principle, the spatial contiguity principle, the temporal contiguity principle, the modified redundancy principle, the signaling principle, the voice principle, the multimedia principle and the image principle were verified to be followed. As discussed in Chapter 2, the reasons why the above multimedia learning principles were followed during the design of the multimedia lesson were because:

1. They met the conditions for the original modality and redundancy principles to apply.
2. They represented different types of multimedia lessons.

3. They ensured that the multimedia lesson was conduciveness to learning.

Statistical analyses in both studies indicated that input modes had no influence on ESL students' learning outcomes when a multimedia lesson was ensured to be an effective one. In other words, the modality and redundancy effects became insignificant for ESL students' learning when the other multimedia learning principles were followed.

Study 2 replicated and extended Study 1 by addressing its limitations. Study 2 achieved the same results as Study 1, which gives greater validity to the findings. In addition, Study 1 and Study 2 support the findings of Mayer et al. (2014) and Van der Zee et al. (2017). Both studies (Mayer et al., 2014; Van der Zee et al., 2017) focused on examining the role of added on-screen text for ESL students' content learning text by comparing learning outcomes from the input mode of graphics +audio and that of graphics + audio + text. They both indicated that redundant text did not have an impact on ESL students' content knowledge learning, which was consistent with the finding of Study 2.

Discussions of Findings

Based on the test results of both Study 1 and Study 2, the reason accounting for the no differences in both retention and vocabulary scores was explained, followed by a discussion about ESL students' no differences in learning outcomes across the three conditions.

Vocabulary and Retention Tests

Just as knowledge retention, ESL students' vocabulary learning was found to have been little influenced by input modes, because both the vocabulary test and the retention test can be regarded as one type of test: the retention test.

The vocabulary test focused on assessing how well the meanings of the technical terms were retained. It did not address other aspects of vocabulary learning, such as pronunciation and

use. Both the vocabulary test and the retention test focused on assessing the retention of key facts explicitly taught in the multimedia lesson and can therefore be regarded as one type of test, which is the retention test.

There is also statistical evidence in both Study 1 and Study 2 supporting that both retention and vocabulary tests might have measured retention. According to Eisinga, Grotenhuis, and Pelzer (2013), standardized Cronbach's alpha was appropriate to examine the reliability for two items measure for one construct. In Study 2, the standardized coefficient alpha ($\alpha = 0.75$) indirectly supported that the retention test and the vocabulary test might have measured a single construct of knowledge retention. The retention and the vocabulary tests respectively assessed ESL students' retention of the five steps of lightning and the meanings of five key concepts. This can account for the no significant differences in both retention and vocabulary scores across the three groups.

Explanations of Findings

Test results of both Study 1 and Study 2 indicated that ESL students learned the same from the three input modes of graphics + audio, graphics + text, and graphics + audio + text. This might be because ESL students ignored verbal inputs of on-screen text, narration, or both and depended on graphics and added keywords to learn, which was explained as follows.

Verbal input was ignored. Based on the test results of both Study 1 and Study 2, ESL students' learning via the graphics + audio was the same as that via graphics + text. ESL participants might have difficulty accessing new information delivered by L2 English audio or text (Cheng et al., 2004; Huang et al., 2016), while they had no issue accessing the same information delivered by graphics, especially when graphics in the multimedia lesson was effective enough to teach the new content due to the control of the multimedia lesson. Based on

the cognitive load theory (Sweller, 2014), when too much information of both graphics and words is presented to ESL students at the same time, their working memory can become overloaded and much of the new information will be lost. Therefore, to reduce the cognitive load for more efficient learning, ESL students might barely use L2 English text or audio that they had difficulty accessing, which could explain the no difference in learning outcomes between the graphics + audio and graphics + text groups.

Similarly, the reason why the redundancy principle did not apply might also be that ESL participants in the two groups ignored audio and audio + text and relied on graphics to reduce the cognitive load for more efficient learning. In this way, learning results from graphics + audio and those from graphics + audio + text were the same.

Therefore, in both Study 1 and Study 2, the no difference in learning outcomes across the three groups might be due to effective graphics resulting from the carefully controlled multimedia lesson. When graphics in the multimedia lesson was controlled to be sufficient enough to teach new content knowledge alone, ESL students could ignore English words (both textual and auditory) that they have difficulty accessing and depend on graphics to learn. Such an assumption could have been validated if there had been another input mode consisting of only graphics in the experiment. If ESL students still learned the same from the input mode of graphics as from the other three input modes, then it would be more certain that ESL students indeed ignored verbal inputs and learned solely by graphics.

Added keyword text in the multimedia lesson. As discussed above, both on-screen text and narration that convey the same verbal messages to teach the content knowledge could have been ignored by ESL students to account for the no differences in learning outcomes across the three conditions. This explanation was supported by the presence of some keywords in the

multimedia lesson. In both Study 1 and Study 2, the pre-training principle (Mayer, 2009) and modified redundancy principle (Mayer & Johnson, 2008) added keywords to the multimedia lesson. Different from text that consists of complete sentences, keywords provided by the pre-training and modified redundancy principles summarize text and are placed to their corresponding graphics. There is a possibility that ESL students relied on neither English audio nor its verbatim on-screen text to learn when the pre-training principle and modified redundancy principle ensured some key verbal messages to be accessed.

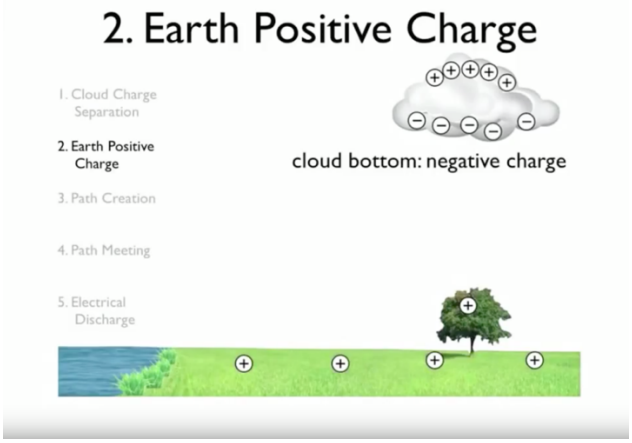

Figure 6. A screenshot from the multimedia lesson that followed the modified redundancy principle.

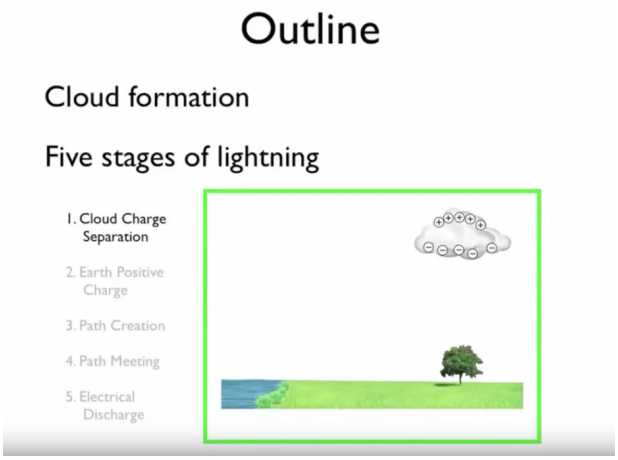

Figure 7. A screenshot from the multimedia lesson that followed the pre-training principle.

Specifically, the combination of the modified redundancy principle (i.e. placing critical concepts near their corresponding graphics directed learner's attention to the concepts as illustrated by a snapshot of the multimedia lesson shown in Figure 6.) and the pre-training principle (i.e. critical concepts were listed at the beginning of the multimedia lesson as illustrated by a snapshot of the multimedia lesson shown in Figure 7.) ensured that some critical on-screen keyword text, independent of audio and its verbatim on-screen text, was added to the multimedia lesson.

Such keyword text is both critical and facilitative for learning, because it "helps manage essential processing (by guiding the learner's attention to the keywords in the narration and the key action in the graphic) while not adding to extraneous processing (by presenting only a few words and placing them next to the portion of graphic they describe)" (Mayer & Johnson, 2008, p. 385). Therefore, ESL learners might have ignored the more complicated L2 English audio and text, and resorted to such keyword text for more efficient learning. If this were so, the participants in the three groups might have learned the same amount of key verbal information that the pre-training and modified redundancy principles enabled, which can account for no differences in learning outcomes across the three input conditions.

This assumption was supported by the vocabulary test results. In both Study 1 and Study 2, participants in the three conditions, including those in the graphics + audio group who did not have access to textual input, achieved the same learning results. For this to happen, participants in the graphics + audio group must have been able to recognize the technical term first, and then give its explanation as those in the other conditions did. The added keyword text ensured that participants in the graphics + audio group could also have access to the technical terms the same way participants in the other conditions did.

To summarize, ESL students had difficulty accessing verbal messages due to their lack of L2 English proficiency (Cheng et al., 2004; Huang et al., 2016), but their accessing graphics was not influenced by lack of L2 language proficiency. Therefore, to reduce the cognitive load in the working memory for more efficient learning, ESL students in the three input conditions had ignored audio and on-screen text, and depended on graphics and added keyword text to learn, which can account for no differences in learning outcomes across the three input conditions. An interview or a survey focusing on the ways participants of the three groups watched the multimedia lesson are justified. For example, a follow-up qualitative study can complement Study 1 and Study 2 for more accurate understanding of how input modes influence ESL students' multimedia learning, as well as how they use verbal inputs in a multimedia lesson.

Strengths and Limitations

Strengths of Study 1 and Study 2

Both Study 1 and Study 2 compared ESL students' content knowledge learning via three input modes in the modality and redundancy principles. Study 2 extended Study 1 by addressing its limitations and achieved the same results, which gives greater validity to the findings.

Empirical contribution. The empirical contribution is that both the modality and redundancy principles disappeared for ESL students who received multimedia lessons in L2 English—that is, students learned the same via graphics + audio, graphics + text, and graphics + audio + text. Additionally, since ESL students usually come to the US universities to learn content knowledge in specific academic fields, both Study 1 and Study 2 compared the three input modes in terms of their influences on ESL students' content knowledge learning, rather than merely language development, which enriched the limited literature. Last, both Study 1 and Study 2 addressed two common internal validity threats in previous studies brought about by

uncontrolled multimedia materials and non-translated instruments, which enabled a more rigorous investigation into the research questions.

Theoretical contribution. The finding that ESL students learned the same from the three input modes of graphics + audio, graphics + text, and graphics + audio + text lends support to the cognitive load theory (Sweller, 2014) as well as the cognitive theory of multimedia learning (Mayer, 2009). To explain no differences in learning outcomes across the three conditions, ESL students could have chosen to give little attention to verbal input of narration, its verbatim text or both, and learn from graphics and added keywords in order not to overload visual channel, which also exemplifies reducing the cognitive load in the working memory for more effective learning. Additionally, Study 1 and Study 2 enriched the understanding of the modality and redundancy principles by including a discussion focusing on non-native learners' learning. The finding that the modality and redundancy principles disappeared when other principles were followed could also spark discussions about the interactions between different learning principles.

Practical contribution. Both Study 1 and Study 2 indicated that input modes did not have an impact on ESL students' content knowledge learning, which can inform content area professors to design a class teaching the same content via different input modes to meet different learning preferences, learning situations, and instructional resources.

The finding that ESL students' learning was not influenced by the input mode makes many empirical applications possible. For example, for cost-effectiveness, there is no need to provide additional captioning. Since the input modes of graphics + text and that of graphics + audio can lead to the same learning results, visual learners might prefer to learn via the mode of graphics + text in a quiet learning environment, which exemplifies different learning options made possible by the findings of Study 1 and Study 2.

Contribution to the media debate. Both Study 1 and Study 2 attempted to explore what Kozma (1994) believed key in the media debate: if there was a relationship between media and learning. However, the finding that ESL students' learning was not influenced by input modes provides evidence to several tenets of Clark's (1994) position in the debate. The same learning results via three input modes of graphics + audio, graphics + text, and graphics + audio + text support Clark's (1994) fundamental position in the great media debate that media do not influence learning. It also exemplifies Clark's (1994) replaceability test implying that a medium or set of media attributes (i.e. input modes in this case) can always be replaced by traditional methods, such as lectures. Additionally, since the input mode of graphics + audio + text and that of graphics + audio could produce the same learning result, it is more cost effective not to provide ESL students with added captioning in a lecture, which supports Clark's (1994) argument that the decision to choose certain medium or media attributes is mostly economic.

Limitations of Study 2

Although Study 2 has addressed the limitations as well as the implementational issues of Study 1, Study 2 is still limited by some validity threats.

First, although participants were expected to be from a more representative population speaking different L1s, the recruited ESL participants in Study 2 turned out to be a rather homogenous sample, composed of a great majority of L1 Chinese speakers. This limits the generalizability of the findings in relation to speakers of other L1s, and therefore creates a threat to external validity. There is little empirical evidence yet showing that different L1s would yield to different learning outcomes from the three input modes, which calls for future empirical studies to substantiate such a claim.

Second, Study 2, as other studies examining similar questions, only assessed short-term retention and transfer right after the treatment, while it is of greater empirical significance if long-term knowledge retention and transfer had been examined. For example, the same participants could have been tested again after one month to compare how much content knowledge was still retained across the three groups. Additionally, this experimental study was conducted in an artificial setting, and conducting it in an actual classroom situation would allow for follow-up assessments on long-term knowledge retention and transfer. Therefore, lack of assessment on long-term learning and the artificial setting both limit the generalization of the findings to other contexts and create external validity threats.

Third, there might not be sufficient number of participants in each group. In Study 2 there were 20 participants in each of the three groups, more than those utilized by Mayer et al. (2014) and Study 1 that also examined different input modes for ESL students' learning. However, Mayer (2011) suggested that at least 25 participants should be recruited in each group to compare learning outcomes from different treatments.

Next, Study 2 did not examine the possible interaction effect between English proficiency-levels in each group and learning outcomes. Although TOEFL reading and listening scores were collected in the questionnaire, the reading and listening scores of all participants fell into the category of *High* (22-30) in relation to *Intermediate* (15-21) and *Low* (0-14) on TOEFL score interpretation. In other words, the ESL students were considered at high proficiency level based on their TOEFL listening and reading scores, which might be because the university usually requires minimum TOEFL scores (i.e. 22 or above) for international students to be admitted. TOEFL scores in Study 2 were insufficient to categorize ESL students into groups of different proficiency levels. Therefore, it is possible that the interventions worked differently for

ESL students of different proficiency levels, which is a threat to external validity. Further studies addressing these limitations are called for to yield more rigorous answers to the research questions.

Last, the learning time was short and lasted for only five minutes. Some ESL students might need some time to adjust their assigned input mode for effective learning to happen. Since the multimedia lesson was short and information-loaded, they might have lost a great chuck of information before learning really happened, which is threat to both internal and external validities.

Conclusion

Overview of the Study

Study 1 and Study 2 both compared the three input modes in the modality and redundancy principles for ESL students' content knowledge and technical vocabulary learning, and simultaneously tested the applicability of the two principles for ESL students. The three input modes were graphics + audio, graphics + text, and graphics + audio + text. Study 1 suffered from some shortcomings. Study 2 replicated and extended Study 1 by addressing its shortcomings and implementation issues for more rigorous answers to the research questions.

In Study 1 and Study 2, common validity issues brought about by uncontrolled multimedia materials in related studies were examined. Then, the researcher utilized 10 multimedia-learning principles to control for the multimedia lesson to ensure that 1. It was conducive to learning; 2. It was representative of other multimedia lessons; 3. The conditions for the modality and redundancy principles were met. Additionally, since ESL students' lack of language proficiency might undermine their abilities to understand test questions and produce answers, L1 translations were provided in the instruments and L1s were allowed to answer

questions, which led to more accurate assessment of content knowledge learning.

Both Study 1 and Study 2 both indicated that input modes did not have an impact on ESL students' learning. Consequently, the modality and redundancy principles became insignificant for ESL students' learning when a combination of the coherence principle, the pre-training principle, the spatial contiguity principle, the temporal contiguity principle, the modified redundancy principle, the signaling principle, the voice principle, the multimedia principle and the image principle was followed in the multimedia lesson to ensure its conduciveness to learning. The no differences in learning outcomes across the three input mode groups in Study 2 align with those of Study 1, which gives greater validity to the findings.

Such no differences in learning across the three groups might be because participants ignored narration, text, or both and depended on graphics and the added keywords to reduce cognitive load in the working memory for more efficient learning. The findings of Study 1 and Study 2 extended the original modality and redundancy principles by examining their applicability for ESL population and discussing their conditions. The findings also provided empirical evidence for designing effective multimedia lessons and providing multiple learning options to better serve the ESL population on university campuses.

Significance

Technological advances enable multimedia instruction to be more accessible and popular in classrooms. Multimedia learning principles should ensure effective learning of students that consist of not only native speakers of the instructional language, but also those who do not speak the instructional language as their native language, such as the ESL students in American universities. However, the modality and redundancy principles (Mayer, 2009) failed to address

this population. This study attempted to fill this gap to examine the applicability of the modality and redundancy principles to ESL students' learning.

Theoretically, the findings of the study extended the modality and redundancy principles by including ESL students in the discussion. Additionally, they provided more evidence for both the cognitive theory of multimedia learning and the cognitive load theory by investigating the role of verbatim on-screen text in ESL students' learning. Empirically, the findings of this study enriched the understanding of ESL students' learning through multimedia, and informed both ESL instructors and content-area professors of how to take advantage of multimedia input modes to optimize their ESL students' learning. The findings also provided insights to design inclusive multimedia instruction that is accessible for both native English-speaking students and ESL students.

Suggestions for Further Studies

Some possible paths of investigation stem from Study 1 and Study 2, especially from the limitations of Study 2.

First, based on the findings of Study 2, the modality and redundancy effects became insignificant for ESL students when the combination of ten principles was followed to convert a raw video into an instruction-appropriate multimedia lesson. Further studies that utilize different combinations (e.g. a combination without the voice and personalization principles) of multimedia learning principles to control for multimedia materials are justified to provide more comprehensive answers to the research questions.

Secondly, based on the explanations for findings, a survey study that focused on eliciting from participants how they utilized graphics, audio, and text during their learning is able to generate accurate explanations for the findings. Another survey could focus on examining the

motivation and self-efficacy of the ESL students when they view the same multimedia lesson in different input modes. The answers could be of great empirical importance to design effective multimedia lessons.

Thirdly, as discussed in the limitation section, posttests in both Study 1 and Study 2 were administered right after the experiment and long-term learning was not assessed. Delayed posttest results could be equally important, if not more, for the design of an effective multimedia lesson. Therefore, future studies featuring delayed posttests are justified to compare ESL students' long-term learning from different input modes. Additionally, the modality and redundancy principles were based on a plethora of studies examining immediate learning from the three input modes. Empirical evidence about long-term retention and transfer of knowledge can greatly extend the modality and redundancy principles.

Fourthly, the recruited ESL participants in both Study 1 and Study 2 consisted of a great majority of L1 Chinese speakers of high English proficiency. Since there is empirical evidence that ESL students' proficiency level could be a variable affecting their test performances (Mohsen, 2016; Van der Zee et al., 2017), studies utilizing samples more representative of the demographics and proficiency levels of ESL students on university campuses might be able to yield to more rigorous answers to the research questions.

Next, in Study 1 and Study 2, the modality and redundancy principles were found not to have applied on the condition that ESL students learned from an effective multimedia lesson controlled by the combination of multimedia learning principles. However, the modality and redundancy principles were based on a series of empirical studies using different and uncontrolled multimedia lessons, which threatened both internal and external validities (as explained in Chapter 2). A multimedia lesson not ensured effectiveness for learning, for example,

a multimedia lesson with dyssynchronous pictures and words, is inappropriate to be utilized to investigate learning via input modes, because it should not be used for instruction in the first place. Therefore, it is justifiable to reexamine the modality and redundancy principles when native English-speaking students learn from a multimedia lesson controlled to be conducive to learning.

Appendix A

Transcription of the Multimedia Lesson

What is lightning? A lightening can be defined as the neutralization of an electrical charge separation between the cloud and the ground. And before we are able to see a bolt of lightning, however, many steps occur behind the scenes. In this presentation, I will explain these steps. First, I will explain how cloud formation begins the process of lightning. Then I will explain the five stages of lightning that cumulate with a lightning bolt.

Now in the first stage, electrical charges separate within the storm cloud. In the second stage, the earth acquires a positive charge. In the third stage, two conductive paths are created. In the fourth stage, these paths meet. And in the fifth and final stage, the electricity is discharged in the form of a lightning bolt. The process that underlines lightning begins with cloud formation. When the surface of the earth is warm, moist air near the earth surface becomes heated and rises. As it rises, the air cools, which produces an updraft. As the air in an updraft cools, water vapor condenses into water droplets and forms a cloud. At high altitude, the air temperature is well below freezing, so the upper portion of the cloud is composed of tiny ice crystals. Eventually, the rain drops and ice crystals get large enough that they fall through the cloud driving some of the air from the cloud downward, which produces downdrafts. When downdrafts strike the ground, they spread out in all directions, which is the wind often felt before a lightning storm. These opposing updrafts and downdrafts create the first stage of lightning, which is called Cloud Charge Separation. Positive and negative charges within the cloud are separated with the positive charges at the top and the negative charges on the bottom.

In the next stage, the earth acquires a positive charge. This happens when the cloud charge separation is so strong that the strong negative charge on the bottom of the cloud causes the earth surface to acquire a strong positive charge in the same way the opposite poles of a magnet attract.

In the third stage, of lightning, two conductive paths between the cloud and the ground are created: one headed from the cloud to the ground, and one from the ground to the cloud. The path that begins the cloud is called the Stepped Leader, and is created by negative charges heading towards the positively charged ground. As the stepped leader approaches the earth, charges on the surface begin responding with a conductive path of their own, called a Streamer. That is a positively charged streamer.

The fourth stage occurs after the negatively charged stepped leader and the positively charged streamer meet. The two paths have completed their journey. With these paths complete, current will be able to flow between the earth and the cloud.

And the fifth and final stage is what we see as lightning. The strike is a sudden massive flow of an electrical current, moving from the cloud to the ground in order to neutralize the separation. And because there is an enormous amount of current, there is also an enormous amount of heat. This heated air causes the brilliant white blue flash that we see.

Appendix B

Instruments with Chinese Translations

Demographics Questionnaire

1. Gender:
2. Age:
3. Native language:
4. Months in the US:
5. Year in college:
6. Years of English learning at school:
7. Your SAT score:
8. TOEFL reading score:
9. TOEFL listening score:

Prior Knowledge Survey

Please circle the number of the sentence that applies to you.
请圈上您同意句子前的数字。

1. I regularly read the weather maps in the newspaper.
2. I know what a *cold front* means.
3. I know what *cumulous*（积云）and *nimbus*（乱云） mean.
4. I know what a *low-pressure system* means.
5. I can explain what makes the wind blow.
6. I know what this symbol means: 
7. I know what this symbol means: 
8. I can explain how lightning works.

Please rate your knowledge of weather and circle the number of the item that applies to you.
请圈上对应您所掌握气象知识前的数字。

1. Very little
2. Between very little and average(很少和中等之间)
3. Average
4. Between average and very much(中等和很多之间)
5. Very much

Please use English and/or Chinese to answer the following questions.

Based on the information in the video, please write down an explanation of how lightning works and briefly illustrate how lightning works on the diagram.

请根据视频中的信息 1. 写出闪电形成的过程; 2. 在下面的图上简要标明闪电形成的过程。

Steps	**Names**	**Explanations**
1		
2		
3		
4		
5		

Please use English and/or Chinese to answer the following questions.

Based on the information in the video, please explain the following terms.
根据视频中的信息，请定义下面的单词。

Terms	**Explanations**
Updraft	
Downdraft	
Stepped-leader	
Streamer	
Neutralization	

Please use English and/or Chinese to answer the following questions.

Based on what you have learned from the video, please answer the following questions.
根据视频中的信息，请回答下面的问题。

(a) Before lightning, is an airplane flying on top of the clouds positively or negatively charged? And why?
在闪电形成前，在云层上飞行的飞机是带正电还是负电，为什么？

(b) When you are on an open area and see lightning in the near distance, is it safe to hide in a ditch? If it is not, what should you do?
当你在一大片空地上行走，并看到远处有闪电时，选择躲在附近的沟渠里安全吗？如果不安全，那你会怎样做？

(c) Why do we sometimes see heavy clouds in the sky but no lightning?
为什么我们有时看到天上有密云，但没有闪电？

References

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science, 35*(6), 499–517.

Aldera, A. S.& Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education, 68*, 60–75.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman.

Bellini, J. L., & Rumrill, P. D., Jr. (2009). *Research in rehabilitation counseling: A guide to design, methodology, and utilization* (2nd ed.). Springfield, Ill: Charles C. Thomas.

Bird, S.A. & Williams, J.N. (2002). The effect of bimodal input on implicit and explicit memory: an investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23* (4), 509–533.

Borgh, K., & Dickson, W. P. (1992). The effects on children's writing of adding speech synthesis to a word processor. *Journal of Research on Computing in Education, 24*(4), 533–544.

Bowe, F. G., & Kaufman, A. (2001). Captioned media: Teacher perceptions of potential value for students with no hearing impairments: A national survey of special educators. Spartanburg, SC: Described and Captioned Media Program.

Brown, C. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin & M. Haynes & J. Coady (Eds.), *Second language reading and*

vocabulary learning. Norwood, N.J.: Ablex.

Butler, S., Urrutia, K., Buenger, A., Gonzalez, N., Hunt, M., & Eisenhart, C. (2010). A review of the current research on vocabulary instruction. *United States Department of Education, National Reading Technical Assistance Center. Washington, DC: Authors.*

Casado, M.A. & Dereshiwsky M. I. (2001). Foreign language anxiety of university students. *College Student Journal, 35*(4), 539–552.

Cheng, L., Myles, J., & Curtis, A. (2004). Targeting Language Support for Non-Native English-Speaking Graduate Students at a Canadian University. *TESL Canada Journal*, *21*(2), 50 - 71. https://doi.org/10.18806/tesl.v21i2.174

Chung, J.(1999). The effects of using video text supported with advance organizers and caption on Chinese college students' listening comprehension: an Empirical Study. *Foreign Language Annals, 32* (3), 295–308.

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning.* San Francisco, CA: Pfeiffer.

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research, 53*(4), 445-459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(1), 21-29.

Diamond, L. & Gutlohn, L. (2006). *Vocabulary Handbook*. CORE Literacy Library: Berkeley, California.

Doty, D. E., Popplewell, S. R., & Byers, G. O. (2001). Interactive CD-ROM storybooks and young readers' reading comprehension. *Journal of Research on Computing in Education, 33*(4), 374-384.

Eisinga, R. N., Grotenhuis, H. F. t., & Pelzer, B. J. (2013). The reliability of a two-item scale: Pearson, cronbach or spearman-brown? *International Journal of Public Health, 58*(4), 637-642. doi:10.1007/s00038-012-0416-3George, D., & Mallery, P. (2009). *SPSS for windows step by step: A simple guide and reference, 16.0 update* (9th ed.). Boston, MA: Allyn and Bacon.

Goldenberg, L. B., Heinze, J., & Ba, H. (2004). *What middle grade students say about learning science with multimedia*. Presented at the National Educational Computing Conference, New Orleans, LA.

Granoff, S., & Whiting, J. (2010). The effects of multimedia input on comprehension of a short story. *TESL-EJ: Teaching English as a Second or Foreign Language, 14*(2).

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85–93.

Hayati, A. & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42*, 1, 181–192.

Hernandez, S. S. (2004). *The effects of video and captioned text and the influence of verbal and spatial abilities on second language listening comprehension in a multimedia environment*. New York University, New York: Unpublished doctoral dissertation.

Hiebert, E. H., & Kamil, M. L. (2005). *Teaching and learning vocabulary: Bringing research to practice*. Mahwah, N.J: L. Erlbaum Associates.

Hinkel, E. (2005). *Handbook of research in second language teaching and learning*. Mahwah, N.J: L. Erlbaum Associates.

Hirai, A. (1999). The relationship between listening and reading rates of Japanese ESL learners. *Modern Language Journal, 83,* 367–384.

Huang, H. C. & Eskey, D. E. (1999). The effects of closed-captioned television on the listening comprehension of intermediate English as a foreign language (ESL) students. *Journal of Educational Technology Systems, 28*(1), 75–96.

Huang, Y.M., Shadiev, R., & Hwang, W.Y. (2016). Investigating the effectiveness of speech-to-text recognition applications on learning performance and cognitive load. *Computers & Education, 101*, 15-28.

Inceçay, V., & Koçoglu, Z. (2016). Investigating the effects of multimedia input modality on L2 listening skills of Turkish EFL learners. *Education and Information Technologies*, 1–16.

Joe, A., Nation, P., & Newton, J. (1996). Vocabulary learning and speaking activities. *English Teaching Forum, 34*(1), 2–7.

Jung, U. (1990). The challenge of broadcast videotext to applied linguistics. *IRAL, 28*(3), 201–220.

Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing, 25*(2), 347-373. doi:10.1007/s11145-010-9272-9

Koolstra, C. M., & Johannes W. J. Beentjes. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development, 47*(1), 51–60.

Kozan, K., Erçetin, G., & Richardson, J. C. (2015). Input modality and working memory: Effects on second language text comprehension in a multimedia learning environment. *System, 55*, 63–73.

Kozma, R. B. (1991). Learning with Media. *Review of Educational Research, 61*(2), 179-211.

Kozma, R. B. (1994). The Influence of Media on Learning: The Debate Continues. *School Library Media Research SLMQ 22*(4).

Lambert, W. (1986). Pairing first- and second- language speech and writing in ways that aid language acquisition. In J. Vaid (Ed.), *Language processing in bilinguals: Psychological and neurological perspectives*. Hillsdale NJ: Lawrence Erlbaum.

Lee, H., & Mayer, R. E. (2018). Fostering learning from instructional video in a second language. *Applied Cognitive Psychology, 32*(5), 648-654. doi:10.1002/acp.3436

Lightbown, P. & Spada, N. (2013). *How Languages Are Learned*. Oxford: Oxford University Press.

Lin, J. J. H., Lee, Y. H., Wang, D. Y., & Lin, S. S. J. (2016). Reading Subtitles and Taking Enotes While Learning Scientific Materials in a Multimedia Environment: Cognitive Load Perspectives on EFL Students. *Educational Technology & Society, 19* (4), 47–58.

Linebarger, D. L. (2001). Learning to read from television: The effects of using captions and narration. *Journal of Educational Psychology, 93,* 288–298.

Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers and Education*. *217*, 190–200. doi.org/10.1016/j.compedu.2018.08.025

Liversidge, G. B. (2000). *The role of closed captioning in second language acquisition* (Order No. 9965992). Available from ProQuest Dissertations & Theses Global. (304644269). Retrieved from http://search.proquest.com/docview/304644269?accountid=14214

Lund, R. (1991). A comparison of the second language listening and reading comprehension. *Modern Language Journal, 75*, 196–204.

Macaro, E. (2010). *Continuum Companion to Second Language Acquisition*. London: Continuum.

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*(3), 321–328.

Mayer, R.E. (1995). 'Cognitive Theory for Education: What Teachers Need to Know'. In Lambert, N.M. & McCombs, B.L. (Eds.). *How Students Learn—Reforming School Through Learner-Centered Education*. Washington, D.C.: American Psychological Association, 353–378.

Mayer, R. E. (1997). Multimedia learning: are we asking the right questions? *Educational psychologist, 32*, 10-19.

Mayer, R. E. (2001). *Multimedia Learning.* New York: Cambridge University Press.

Mayer, R. E. (2003). Elements of a science of e-learning. *Journal of Educational Computing Research*, *29*(3), 297–313.

Mayer, R. E. (2005a). Cognitive theory of multimedia learning. In R.E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning.* New York: Cambridge University Press.

Mayer, R. E. (2005b). Introduction to multimedia learning. In R.E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning.* New York: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed). New York: Cambridge University Press.

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education*, *44*, 543–549.

Mayer, R.E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2014a). Multimedia Instruction. *Handbook of Research on Educational Communications & Technology,* 385–399.

Mayer, R. E. (2014b). Research-based principles for designing multimedia instruction. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.). *Applying science of learning in education: Infusing psychological science into the curriculum.*

Mayer, R. E. (2014c). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E. (2014d). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning* (pp. 43-71). New York, NY, US: Cambridge University Press.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82,* 715–726.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), The Cambridge handbook of multimedia learning (2nd ed.) (pp. 279–315). New York: Cambridge University Press.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology, 28*(5), 653–660.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*(2), 380–386.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90,* 312–320.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational*

Psychology, *91*(4), 638–643.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*, 107–110.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Signaling, pre-training, and modality. In R. E. Mayer (Ed.), The Cambridge handbook of multimedia learning (2nd ed.) (pp. 316–344). New York: Cambridge University Press.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86,* 389–401.

Meskill, C. (1996). Listening skills development through multimedia. *Journal of Educational Multimedia and Hypermedia, 5*(2), 179–201.

Miller, M. V. (2011). A system for integrating online multimedia into college curriculum. *Journal of Online Learning and Teaching, 7*(2), 294.

Mohsen, M. A. (2016). The use of help options in multimedia listening environment to aid language learning: A review. *British Journal of Educational Technology, 47*(6), 1232–1242.

Montero Perez, M. & Desmet, P. (2012). The effect of input enhancement in L2 listening on incidental vocabulary learning: A review. *Procedia - Social and Behavioral Sciences, 34*, 153–157.

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2

listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94,* 156–163.

Moulton, S. (2010, July 13). *How lightning works* [Video file]. Retrieved from https://www.youtube.com/watch?v=_bY-8ozney8

Nation, I. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nelson, J., Balass, M., Perfetti, C. (2005). Differences between written and spoken input in learning new words. *Written Language and Literacy, 8*(2), 101–120.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*(1), 95–106.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, and Winston.

Quade, D. (1967). Rank analysis of covariance. Journal of the American Statistical Association, 2(320), 1187–1200.

Rodgers, M. P. H., & Webb, S. (2016). The effects of captions on EFL learners' comprehension of English-language television programs. *Calico Journal, 34*(1), 20–38.

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). Assessing culturally and linguistically diverse students: A practical guide. New York, NY: Guilford Press.

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal Design for Learning*. Alexandria, VA: Association for Supervision and Curriculum Development

(ASCD).

Rivas, B. K. (2011). *Vocabulary development through shared readings: The potential role of explicit instruction and supplemental multimedia in the preschool ELL classroom.*

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Shadiev, R., Huang, Y.M., & Hwang, J.P. (2017). Investigating the effectiveness of speech-to-text recognition application on learning performance, attention, and meditation. *Educational Technology Research and Development, 65*(5), 1239-1261.

Shatz, M., & Wilkinson, L. C. (2010). *The education of English language learners: Research to practice.* New York: Guilford Press.

Shea, P. (2000). Leveling the playing field: A study of captioned interactive video for second language learning. *Journal of Educational Computing Research, 22* (3), 243–263.

Shin, D. (1998). Using video-taped lectures for testing academic listening proficiency. *International Journal of Listening, 12*, 56–79.

Smidt, E., & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer Assisted Language Learning, 17*(5), 517-556.

Spinelli, C. G. (2008). Addressing the issue of cultural and linguistic diversity and assessment: Informal evaluation measures for English language learners. Reading & Writing Quarterly, 24(1), 101–118.

Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*, 661–699.

Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer

(Ed.), The Cambridge handbook of multimedia learning (2nd ed.) (pp. 27–42). New York: Cambridge University Press.

Sweller, J., Ayres, P. & Kalyuga, S. (2011). *Cognitive load theory*. NY: Springer.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*(3), 422–427.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*(4), 257–287.

Toh, S. C., Munassar, W. A. S., & Yahaya, W. A. J. W. (2010). Redundancy effect in multimedia learning: A closer look. Curriculum, technology and transformation for an unknown future. In *Proceedings Ascilite Sydney* (pp. 988–998).

Tseng, W., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modelling approach. *Language Learning, 58*, 357–400.

Tsou, W., Wang, W. & Li, H.Y. (2002). How computers facilitate English foreign language learners acquire English abstract words. *Computers & Education, 39*(4), 415–428.

Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing, 27*(4), 493–513.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*, 46–65.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology: A Refereed Journal for Second and Foreign Language Educators, 14*(1), 65-86.

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition, 23*(3), 345–368.

Yoon, T. (2013). Exploring a literacy development in young Korean ESL students with online E-books. *Journal of Language Teaching and Research, 4*(2), 253–259.

YINAN LIU

108 Walnut St, Fayetteville, NY 13066 | 315-459-1106 | yliu15@syr.edu | www.esolvideos.com

EDUCATION

Syracuse University, Syracuse, NY

Ph.D. in Literacy Education 2019

Syracuse University, Syracuse, NY

M.A. in Linguistic Studies Concentration: Second Language Acquisition 2009

Dalian University of Foreign Languages, Dalian, Liaoning, China

B.A. in English Minor: Information Studies 2003

AWARDS AND CERTIFICATES

Doctoral tuition scholarship, Syracuse University 2012-2017

Master's tuition scholarship, Syracuse University 2007-2009

Certificate of University Teaching, Syracuse University 2009

University Teaching Qualifications Certificate, Chinese Ministry of Education 2005

PATENTS

A Language Teaching System Based on Movies (CN2019109187957) 2019

An Auditory System for Language Teaching (CN 2019108815060) 2019

Adaptive Multimedia Language Learning System (US 62749749) 2018

TEACHING EXPERIENCE

Syracuse University, Syracuse, NY

Teaching Assistant 2012-2017

•Developed the syllabus and overall course structure for CLS 105: College Learning Strategies, assisted in the course, met with individual students on a weekly basis, provided students with mentoring, and graded assignments

TOP IELTS School (www.topielts.com), Dalian, Liaoning, China

Co-founder and Instructor 2009-2012

•Recruited ESL teachers, selected textbooks, designed and evaluated courses, taught TOEFL and IELTS classes, as well as engaged in various administrative obligations

Dongbei University of Finance and Economics, Dalian, Liaoning, China

Tenure-tracked Lecturer 2009-2012

•Taught LIN 601: Introduction to Linguistics, ENG 401: English Reading, ENG 404: English Communications, and ENG 503: IELTS Reading and Speaking

Syracuse University, Syracuse, NY

Part-time Instructor 2007-2009

•Taught Chinese 101, 102, 201, 202 and 301

Dongbei University of Finance and Economics, Dalian, Liaoning, China

Assistant Professor 2003-2007

•Taught ENG 401: English Reading and ENG 404: English Communications

PROFESSIONAL SERVICES AND AFFILIATIONS

Reviewer for the Asia-Pacific Educational Researcher 2019-present

Reviewer for Journal of Adolescent and Adult Literacy 2016-present

Member of International Literacy Association 2015-present

Member of American Educational Research Association 2015-present

Member of Teaching English as a Foreign Language 2015-present

PUBLICATIONS

Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers and Education*. *217*,190–200.

Liu, Y. (2011). Comparison of real estate naming strategies in China and US. *Chinese and International Entrepreneurs, 364*, 45–56.

Wang, L., Li, J., Qi, L., **Liu, Y.**, & Ma, L. (2007). *How to prepare for IELTS 4.* Beijing, China: Foreign Languages Teaching and Research Press.

Wang, L. & **Liu, Y.** (2005). The patterns of compliment responses in Beijing: A gender perspective. *Foreign Language Teaching and Research. 85*, 44–49.