6-2-2017

Effects of Multimedia Instruction on L2 Acquisition of High-Level, Low-Frequency English Vocabulary Words

Euna Cho

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: http://academicworks.cuny.edu/gc_etds

Part of the Applied Linguistics Commons, Curriculum and Instruction Commons, Educational Methods Commons, First and Second Language Acquisition Commons, Language and Literacy Education Commons, and the Other English Language and Literature Commons

Recommended Citation

Cho, Euna, "Effects of Multimedia Instruction on L2 Acquisition of High-Level, Low-Frequency English Vocabulary Words" (2017). CUNY Academic Works.
http://academicworks.cuny.edu/gc_etds/1986
EFFECTS OF MULTIMEDIA INSTRUCTION ON L2 ACQUISITION OF HIGH-LEVEL, LOW-FREQUENCY ENGLISH VOCABULARY WORDS

by

EUNA CHO

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2017
Effects of Multimedia Instruction on L2 Acquisition of High-Level, Low-Frequency English Vocabulary Words

by

Euna Cho

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

__________________________________________
Date                              Gita Martohardjono
                                 Chair of Examining Committee

__________________________________________
Date                              Gita Martohardjono
                                 Executive Officer

Supervisory Committee:

Gita Martohardjono

Beatriz Lado

Rebecca Curinga

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

Effects of Multimedia Instruction on L2 Acquisition of High-Level, Low-Frequency English Vocabulary Words

by

Euna Cho

Advisor: Professor Gita Martohardjono

The present study examined the effects of multimedia enhancement in video form in addition to textual information on L2 vocabulary instruction for high-level, low-frequency English words among Korean learners of English. Although input-based incidental learning of L2 vocabulary through extensive reading has been conventionally believed to be appropriate for high-frequency words, intentional or explicit vocabulary learning is suggested to be more sensible or realistic for the acquisition of low-frequency academic words. Multimedia support in foreign language instruction has revealed benefits in promoting direct teaching and explicit learning of L2 vocabulary; moreover, adding textual information to video seems to boost students’ understanding of the learning materials. Under the theoretical frameworks such as the dual-coding theory and the cognitive load theory, the study investigated (1) multimedia effects on vocabulary acquisition of advanced-level infrequent words, (2) the best way to offer multimedia by combining the optimal modes of presentation, and (3) the aspects in multimedia support that can help students with acquisition and retention of unfamiliar words.
Seventy-four Korean students who were preparing for the GRE for graduate study in the U.S. participated in the experiment. They were randomly divided into four different groups and were given instruction on 34 GRE vocabulary words in four different conditions: Text-only, Text+Audio, Text+Video, or Text+Audio+Video. After each treatment, immediate post-tests and seven-day delayed post-tests were administered to evaluate participant score change from the pre-tests.

Results indicate that multimedia presentation has a greater positive effect on learning than text-only presentation, supporting the dual-coding theory. Among the types of multimedia support, Text+Audio+Video appears to be better than Text+Audio or Text+Video, suggesting the benefit of the combination of audio and video. Moreover, multimedia audio-visual support was found to be more advantageous when supplemented with a linguistic cue in the form of a precise definition or synonym of an unknown word. Findings have both theoretical and pedagogical implications in L2 vocabulary acquisition of high-level, low-frequency English words in that the study addressed ways to design effective multimedia materials and offered instructional guidelines for multimedia language teaching.
ACKNOWLEDGEMENTS

First and foremost, I would like to express my unqualified gratitude to my advisor and committee chair Dr. Gita Martohardjono. From my first qualifying exam and onward, she has provided me with excellent guidance and has encouraged me to move forward with my research project. I have been inspired by her unwavering mentorship and insightful advisement. I am also grateful to my committee members, Dr. Beatriz Lado and Dr. Rebecca Curinga. Professor Lado has been consistently supportive ever since I wrote a final paper for her class, which became the groundwork for my second qualifying exam and has eventually evolved into the dissertation. She has always been enthusiastic about my research topic and has offered me a tremendous amount of valuable comments that enhanced my dissertation in many ways. I also appreciate the advice and feedback from Professor Curinga whose expertise in TESOL helped me with a more hands-on approach to data analysis. In addition, her suggestions about supplementary studies and direct attention to details enabled me to improve the quality of the dissertation. Throughout the dissertation phase, I have felt extremely lucky to have such caring and devoted professors on the committee. Without their continuous support and encouragement, the dissertation could not have been accomplished.

My appreciation also goes to Dr. Elaine Klein, Dr. Martin Chodorow, and Dr. Marcel den Dikken. Professor Klein gave me thoughtful advice and extensive feedback on my second qualifying paper. Her expertise in L2 vocabulary acquisition gave me a deeper understanding of theoretical aspect of vocabulary and guided my literature review section to the next level. Professor Chodorow granted me helpful suggestions and valuable comments on statistical analysis of the research. Professor den Dikken, although
not directly involved in my dissertation, greatly contributed to my first qualifying exam by offering me insightful guidelines about syntactic theories and analyses in East Asian languages.

I extend my thanks to some of my Graduate Center fellow students as well: Christen N. Madsen II for comprehensive statistical analysis, Jennifer Chard Hamano for outstanding proofreading and editing, Matthew Stuck for meticulous work on data entry, and Chun-Yi Peng for insider information about the procedures of developing and writing a dissertation. I am also thankful to Nishi Bissoondial for her assistance with every stage of the program requirements over the years.

I am truly grateful to my colleagues in Korea who have supported my research endeavors. Jeehoon Jung deserves special mention for his immense help in recruiting participants. I appreciate his patience and understanding all through my graduate career. I also thank my dear friend and co-worker, Euna Filchock for her persistent interest and confidence in my work. The heads of PJ English language institute, Mr. Taewoo Lim and Mr. Kihwan Lee, deserve my sincere thanks for providing me with classroom facilities and generously helping me with publicizing subject enrollment. Likewise, I thank each and every one of my Korean students for being part of the research and serving as greatly valuable assets for my dissertation.

Lastly, I thank my family, especially my mom and my sister, who believed in me and have never lost faith in me in all my pursuits over time, which helped me persevere throughout the entire period of my graduate studies and successfully complete the dissertation.
**Table of Contents**

ABSTRACT .............................................................................................................................. iv

ACKNOWLEDGEMENTS ........................................................................................................ vi

List of Figures .......................................................................................................................... viii

List of Tables ........................................................................................................................... x

I. Introduction .......................................................................................................................... 1

II. Literature Review ............................................................................................................. 4  
   1. Second language (L2) vocabulary development ............................................................... 4  
      1.1. L2 vocabulary knowledge: Knowing a word .............................................................. 4  
      1.2. Measuring lexical knowledge .................................................................................. 7  
      1.3. L2 vocabulary size: How much should a learner know? ........................................... 9  
   2. L2 vocabulary instruction ............................................................................................... 12  
   3. Multimedia learning in L2 vocabulary acquisition ......................................................... 16  
      3.1. Multimedia theories ................................................................................................ 17  
      3.2. Multimedia annotations in computer programs ....................................................... 19  
      3.3. Video effects in multimedia instruction on L2 vocabulary ..................................... 27  
      3.4. Linguistic cues in multimedia presentation ............................................................ 30

III. Pilot Study ....................................................................................................................... 35  
   1. Participants ................................................................................................................. 35  
   2. Experimental design .................................................................................................... 35  
   3. Results ....................................................................................................................... 40  
   4. Discussion and conclusions ....................................................................................... 44  
   5. Implications from the pilot study .............................................................................. 47

IV. Research Questions and Hypotheses ............................................................................. 48  
   1. Research Questions .................................................................................................... 48  
   2. Hypotheses ............................................................................................................... 48

V. Methodology .................................................................................................................... 50  
   1. Participants ................................................................................................................ 50  
   2. Materials .................................................................................................................... 53  
      2.1. Test Items ................................................................................................................ 53  
      2.2. Video clips .............................................................................................................. 55  
   3. Instruments .............................................................................................................. 56  
      3.1. Pre-test and post-tests ............................................................................................ 56  
      3.2. Vocabulary size test ............................................................................................... 57  
      3.3. Michigan Test of English Language Proficiency ................................................... 58  
      3.4. Questionnaire ........................................................................................................ 58  
   4. Treatments ................................................................................................................. 59  
   5. Procedures ................................................................................................................ 61

VI. Analyses ......................................................................................................................... 63  
   1. Participant analysis .................................................................................................... 63  
      1.1. Age ....................................................................................................................... 63
**List of Figures**

Figure 1. Definition of test items ................................................................. 37
Figure 2. Examples of video clips for the Text+Video group .............................. 38
Figure 3. Handout for the Text-only group ....................................................... 39
Figure 4. Example of post-tests (meaning recall test) ........................................ 39
Figure 5. Means of the tests by instruction with error bars .................................. 40
Figure 6. Mean scores of the pre- and post-tests with error bars .......................... 42
Figure 7. Interaction effects between the tests and the instruction type ................. 43
Figure 8. Word frequency for 34 test items ..................................................... 54
Figure 9. Examples of the Vocabulary Size Test (Korean version) ....................... 58
Figure 10. Screenshots of definitions of target words .......................................... 59
Figure 11. Screenshots of target word in subtitles ............................................. 61
Figure 12. Scatterplot of participants’ percent score and length of residence by group with linear regression line .......................................................... 66
Figure 13. Scatterplot of participants’ percent score and Michigan score by group with linear regression line .......................................................... 67
Figure 14. Scatterplot of participants’ percent score and Vocabulary size by group with linear regression line .......................................................... 67
Figure 15. Scatterplot of participants’ percent score and LOR by group with linear regression line .......................................................... 69
Figure 16. Scatterplot of participants’ percent score and Michigan score by group with linear regression line .......................................................... 70
Figure 17. Scatterplot of participants’ percent score and vocabulary size by group with linear regression line .......................................................... 70
Figure 18. Mean percent difference between the pre-test and post-test 1 for definition questions with 95% confidence interval standard error bars by group ............ 75
Figure 19. Scatterplot of participants’ percent difference between the definition pre-test and post-test 1 and LOR by group with linear regression line .............. 76
Figure 20. Scatterplot of participants’ percent difference between the definition pre-test and post-test 1 and Michigan score by group with linear regression line ...... 77
Figure 21. Scatterplot of participants’ percent difference between the definition pre-test and post-test 1 and vocabulary size by group with linear regression line .... 77
Figure 22. Mean percent difference between the pre-test and post-test 2 for definition questions with 95% confidence interval standard error bars by group ............ 82
Figure 23. Scatterplot of participants’ percent difference between the definition pre-test and post-test 2 and length of residence by group with linear regression line .. 83
Figure 24. Scatterplot of participants’ percent difference between the definition pre-test and post-test two and Michigan score by group with linear regression line .... 84
Figure 25. Scatterplot of participants’ percent difference between the definition pre-test and post-test two and vocabulary size by group with linear regression line .... 84
Figure 26. Mean percent difference between the pre-test and post-test 1 for multiple choice questions with 95% confidence interval standard error bars by group . 88
Figure 27. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test one and LOR by group with linear regression line .......... 90
Figure 28. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 1 and Michigan score by group with linear regression line. 90
Figure 29. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test one and vocabulary size by group with linear regression line .......................................................... 91
Figure 30. Mean percent difference between the pre-test and post-test 2 for multiple choice questions with 95% confidence interval standard error bars by group. 95
Figure 31. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 2 and LOR by group with linear regression line ..................... 97
Figure 32. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 2 and Michigan score by group with linear regression line. 97
Figure 33. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 2 and vocabulary size by group with linear regression line. 98
List of Tables

Table 1. Descriptions of high-, mid-, and low-frequency words ........................................11
Table 2. General descriptions of participants (n= 15).......................................................... 35
Table 3. Experimental design of the pilot study ..................................................................... 36
Table 4. Mean scores and standard deviation of pre-test and post-tests ............................... 40
Table 5. Demographic statistics by group (n= 73)................................................................. 52
Table 6. Experimental design ............................................................................................... 62
Table 7. Mean score and percent by group for definition pre-test (n=34) ............................ 65
Table 8. Pearson’s Correlations between percent score and LOR, Michigan score, and vocabulary size on definition pre-test .......................................................... 66
Table 9. Mean score and percent by group for multiple choice pre-test (n=34) ................. 68
Table 10. Pearson’s Correlations between percent score and LOR, Michigan score, and vocabulary size on the pre-test ................................................................. 69
Table 11. 2 x 2 matrix of instructional conditions and the 4 experimental groups .......... 74
Table 12. Mean score, percent, and percent change by group for definition post-test 1 (n=34) .................................................................................................................. 75
Table 13. Pearson’s Correlations between percent change and LOR, Michigan score, and vocabulary size on definition post-test 1 .......................................................... 76
Table 14. Logistic mixed-effects model of participant improvement on definition post-test 1 ......................................................................................................................... 79
Table 15. Mean score, percent, and percent change by group for definition post-test 2 (n=34) ..................................................................................................................... 81
Table 16. Pearson’s Correlations between percent change and LOR, Michigan score, and vocabulary size on definition post-test 2 ......................................................... 83
Table 17. Logistic mixed-effects model of participant improvement on definition post-test 2 ......................................................................................................................... 85
Table 18. Mean score, percent, and percent change by group for multiple choice post-test 1 (n=34) ............................................................................................................. 88
Table 19. Pearson’s Correlations between percent change and LOR, MTELP score, and vocabulary size on multiple choice post-test 1 ......................................................... 89
Table 20. Logistic mixed-effects model of participant improvement on multiple choice post-test 1 .................................................................................................................. 92
Table 21. Mean score, percent, and percent change by group for multiple choice post-test 2 (n=34) ............................................................................................................. 95
Table 22. Pearson’s Correlations between percent change and LOR, Michigan score, and vocabulary size on multiple choice post-test 2 ......................................................... 96
Table 23. Logistic mixed-effects model of participant improvement on multiple choice post-test 2 .................................................................................................................. 99
Table 24. Mean rating of instruction and usefulness by group ............................................... 102
I. Introduction

Vocabulary is central to language and is of great importance to language learning. Despite its importance, second language (L2) vocabulary teaching has been neglected, mainly because of the belief that vocabulary is learned as a result of extensive reading (Laufer, 2006; Laufer & Nation, 2011). This input-based incidental learning is shown to be effective when learning the most frequent 2,000-3,000 words in the L2 (Nation, 2013). However, more infrequent higher-level words do not appear often enough for an L2 learner to remember, and it is not sensible or realistic to engage in reading hoping to encounter such words. Instead, learners are more likely to benefit from direct teaching and study of vocabulary than incidental learning (Laufer, 2006).

One of the main issues involving vocabulary acquisition is how new words are learned and related questions arise with regard to the role of incidental versus intentional learning (Laufer & Nation, 2011). Input-based incidental learning refers to vocabulary learning that takes place as a chance occurrence during reading, an approach that is conventionally taken in first language (L1) vocabulary acquisition. On the other hand, intentional learning is more deliberately word-focused, and is often referred to as language-focused or *form-focused* instruction (FFI) (Ellis, 2001; Ellis, 2008; Laufer & Nation, 2011), to be described more fully below. In L2 vocabulary acquisition, there is some evidence that word-focused intentional instruction is more effective than mere incidental learning, and that focusing on words in non-communicative tasks may yield

---

1. Two major types of form-focused instruction (FFI) are Focus on Form (FonF) and Focus on Forms (FonFs) (Doughty & Williams, 1998; Laufer, 2006). FonF attends to linguistic aspects during a communicative activity, whereas FonFs teaches discrete linguistic structures in separate lessons. The word-focused instruction in the current study is analogous to FonFs. Nevertheless, as this study does not involve or observe any communicative activities, only explicit vocabulary instruction in isolation will be the matter of interest. Hence, the distinction between FonF and FonFs is of no concern to the current research.
better results than acquisition from implicit input (Cobb, 2007; Elgort, 2011; Ellis, 2001, 2008; Laufer, 2006; Laufer & Nation, 2011).

Foreign language instruction using multimedia aids is known to promote intentional or explicit learning of L2 vocabulary, and positive results have been found when vocabulary is learned with multimedia visual cues such as multimedia annotations (Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Yoshii & Flaitz, 2002), self-instruction computer programs (Kim & Gilman, 2008), or classroom instruction with video (Hanley, Herron, & Cole, 1995; Neuman & Koskinen, 1992). Multimedia or multimodal presentation has been broadly supported by the *dual-coding theory* (Clark & Paivio, 1991) which proposes that visual and verbal information is processed in different systems. The visual system first recognizes the input with the eyes and makes pictorial representations while the verbal system encodes representations using input from the ears. The dual-coding theory suggests that learning can be promoted when learners utilize more than one sensory modality such as visual and verbal cues (Clark & Paivio, 1991; Mayer, 1997). Further, the *cognitive load theory* (Sweller, Van Merrienboer, & Paas, 1998), also endorsed by the dual-coding theory, posits that learners’ cognitive resources can become overloaded because the highly limited processing channels have a limit on how much information can be processed. That is, presenting too many words or pictures that are too complicated may overload the working memory and hinder the verbal and visual processing capacities.

The efficacy of multimedia aids in L2 vocabulary teaching has been mostly focused on beginner and intermediate level vocabulary, yet little attention has been paid to high-level infrequent words that are nevertheless important for advanced learners’ particular
needs, for instance, graduate study. Moreover, recent research has chiefly looked at multimedia effects in computer programs rather than in classroom settings. Based primarily on the dual-coding theory (Clark & Paivio, 1991) and the cognitive load theory (Sweller, Van Merrienboer, & Paas, 1998), the present study investigates the effectiveness of multimedia aids on acquisition of high-level, low-frequency\(^2\) English vocabulary (i.e., words shown on the GRE) by comparing four different types of classroom instruction (one control and three multimedia groups): (1) text only (Text-only, control group); (2) text and audio (Text+Audio, reduced video); (3) text and video (Text+Video, reduced audio); and (4) text, audio, and video (Text+Audio+Video). A pre-test was administered before instruction to establish a baseline. An immediate post-test and a delayed post-test were given to examine retention effects.

The paper is organized as follows: Section II reviews L2 vocabulary acquisition and instruction along with multimedia effects in foreign language classes, Section III illustrates the pilot study, Section IV states research questions and hypotheses, Section V describes the methods and procedures, Section VI demonstrates how to analyze the data, Section VII reports the test results, Section VIII summarizes the results, followed by discussion and conclusions in Section IX.

---

\(^2\) The three main frequency levels are high frequency (1,000-2,000), mid-frequency (3,000-9,000) and low-frequency (10,000 and beyond) (Nation, 2012). More information can be found in Section II.
II. Literature Review

1. Second language (L2) vocabulary development

1.1. L2 vocabulary knowledge: Knowing a word

Lexical knowledge can be classified as either receptive or productive (e.g., Nation, 2013; Laufer & Nation, 2011; Laufer & Paribakht, 1998; Milton, 2009). ‘Receptive’ knowledge relates to comprehending words: receiving language input from listening or reading and trying to understand it. ‘Productive’ knowledge is associated with using words: producing language forms by speaking or writing (Nation, 2013). Nation further draws a distinction between ‘meaning recognition/recall’ for receptive knowledge and ‘form recognition/recall’ for productive knowledge. A general consensus is that word comprehension (reception) does not automatically entail correct use of the word (production). It is also agreed that receptive knowledge is more extensive than productive knowledge, and that receptive knowledge develops faster than productive knowledge does (e.g., Laufer & Paribakht, 1998). Stoddard (1929) was one of the earliest studies that compared receptive and productive L2 vocabulary knowledge. In Stoddard, half of the 328 English speakers learning French studied 50 French words and were asked to translate them into their L1, English (receptive learning). The other half learned the same items and asked to translate from English to French (productive learning). As suggested by Nation (2013) above, L2-L1 translation is considered receptive knowledge since it is ‘meaning recall’ (i.e., understanding and recalling the meaning of L2 words even though the evidence is produced in the L1). On the other hand, L1-L2 translation can be seen as productive knowledge as it is ‘form recall’. A subsequent recall test was administered

3 The terms ‘passive’ and ‘active’ are often used synonymously with ‘receptive’ and ‘productive’ (Nation, 2013).
both receptively (French words for English translation) and productively (English words for French translation), resulting in overall higher performance on the receptive test regardless of the learning condition. Another study conducted by Waring (1997) revealed a similar outcome. Waring had participants tested on the same vocabulary items both receptively and productively, on the same day, the next day, and one week later. His findings indicated that the scores on the receptive test were consistently higher than those on the productive tests, especially for the delayed recall, and that receptive learning took less time than productive learning. Similarly, Bonner (2013) looked into discrepancies between reception and production though not vocabulary. He was particularly interested in how an L2 learner’s morphosyntactic knowledge is reflected in the learner’s perception and production of inflections in English. Bonner’s study indicates that there are asymmetries between the reception and production, attributable to deficiencies in performance. He claims that an L2 learner’s knowledge is not directly represented in his/her performance.

However, reception and production are not completely distinct; rather, they are seen as being on a continuum. For example, Beck, McKeown, and Kucan (2002) suggest a scale of word knowledge for native English speakers: (a) no knowledge; (b) general sense; (c) narrow, context-bound knowledge; (d) limited knowledge; (e) rich, decontextualized knowledge. Although L2 lexical knowledge may not follow such a complex pattern, L2 vocabulary knowledge is also widely understood as a continuum with several levels and dimensions (Laufer & Nation, 2011; Laufer & Paribakht, 1998). Nagy and Scott (2001) also provided a number of complexity measures for word knowledge in children’s vocabulary acquisition including: incrementality (several
dimensions of knowledge), multidimensionality (different types of knowledge), polysemy (multiple meanings), interrelatedness (word knowledge dependent on each other), and heterogeneity (different knowledge depending on a word). More elaborated definitions of lexical knowledge have been proposed by other researchers. For example, according to Nation (1990), knowing a word means having inter-related sub-knowledge about its form, position (grammar, collocations), function (frequency, appropriateness), and meaning (associations). Laufer (1990) proposes a slightly different view of word knowledge that consists of its form (phonology, morphology), its syntactic behavior, its meaning (associations, references), and its relationships with other words. Given all the variations on the definition of lexical knowledge, ‘knowing a word’ in this study can be defined as knowing its meaning, hence receptive knowledge, which is measured by means of ‘meaning recall’ tests (i.e. L2-L1 translation) and ‘meaning recognition’ tests (i.e. multiple choice questions).

Another consideration for examining word knowledge is to discriminate between breadth and depth of such knowledge (Anderson & Freebody, 1981 as cited in Milton, 2009). The breadth of knowledge refers to how many words an L2 learner knows (i.e., vocabulary size), and the depth of knowledge refers to how well an L2 learner knows each word (Milton, 2009), acknowledging that ‘knowing a word’ has many connotations as the previous paragraph illustrated. When measuring vocabulary knowledge, both breadth and depth should be taken into account. Vocabulary breadth can involve not merely productive knowledge, but also the passive recognition of word forms separate from meaning; in other words, a learner can recognize the existence of a word whether or not he/she knows its meaning. Depth of knowledge appears to be even more complex, as
previously described, since it involves other aspects of knowledge. Hence, it is recommended that multiple measures and different tests be used to gain more comprehensive information on a learner’s vocabulary knowledge (Milton, 2009).

1.2. Measuring lexical knowledge

A way of measuring lexical competence is to determine an individual’s breadth of vocabulary knowledge by quantifying total vocabulary size, and in doing so, it seems necessary to define a ‘word’. Obviously, different words like the and person are counted as separate vocabulary items, yet things are not as straightforward when accounting for inflectional and derivational forms such as persons or personalize. For this reason, Milton (2009) underlines that there is no single, simple definition of a word that is used in the creation of tests that measure vocabulary knowledge and learning. Still, in recent vocabulary acquisition studies, two broad conventions are used in counting the number of words: lemmatization and word families (e.g., Milton, 2009). A ‘lemma’ includes a headword and its most frequent inflectional derivations; as with all inflectional morphology, this process does not change the part of speech of the headword. Thus, the lemma of the verb govern would include governs, governed, and governing, but not government (Goulden, Nation, & Read, 1990), which involves the process of a derivational (as opposed to an inflectional) morphological change. Another convention uses a ‘word family’, originally proposed by Bauer and Nation (1993), which includes both inflectional and derivational morphological additions to the base word. In this case, government, governor, governable, or ex-governor are all considered to have the same headword govern, in addition to the inflected words that are already included in a lemma
(i.e., *governs*, *governed*, and *governing*). Consequently, the base form of a word family includes more forms than that of a lemma, and thus vocabulary size under the definition of word family will represent smaller figures than lemmatized word count. According to Milton, the approximate vocabulary size using lemmas can be reached by multiplying the vocabulary size in word families by 1.6. For example, if a learner has a vocabulary size of 1,000 word families, s/he knows approximately 1,600 lemmas. Note also that numbers, proper nouns and names are excluded in creating frequency lists or estimating vocabulary size.

In English, printed school texts are known to include approximately 80,000 word families, and if proper nouns, multiple meanings of words, and idioms are counted, the estimate is 180,000 word families (Anderson & Nagy, 1992). Admittedly, some of the words are much more frequently used than others. For instance, the most common three words in English, *the*, *and* and *a/an*, make up 20% of the corpus. Structural vocabulary like prepositions (*in*, *at*, *of*) and auxiliary verbs (*do*, *will*) are also high-frequency words and labeled as level 0 words. In English, level 0 vocabulary items consist of 150-200 words and are often excluded from the word count. At the other end of the spectrum, there are uncommon words that may show up only a few times in a corpus. In Nation (2013), words have been divided into high-, mid-, and low-frequency words based on the word frequency in the British National Corpus (BNC). The 2,000 high-frequency word families cover 86% of the corpus, the 7,000 mid-frequency word families cover 9% of the corpus, and the 50,000 low-frequency word families comprise 1-2% of the corpus.

Nation and Beglar (2007) state that the word family\(^4\) is more appropriate than the

---

\(^4\) Henceforth, ‘word family’ and ‘word’ may be used interchangeably.
lemmatized account because learners have some control of word formation devices when they pass the minimal proficiency level and thus are able to see the relationships among affixed members of a word family. Accordingly, the current study adopts the word family definition in measuring the breadth of receptive vocabulary knowledge.

1.3. **L2 vocabulary size: How much should a learner know?**

A great amount of research has been implemented towards selecting the words that are most important to L2 English vocabulary learning (Laufer & Nation, 2011). A General Service List (GSL) of English words that contains 1,964 word families was created by Michael West in 1953 and covers 80% of the running words in general English context. In 2013, Browne and his colleagues released an updated list, a New General Service List (NGSL), in an effort to increase generalizability and validity of the existing list (Browne, 2014). NGSL includes the 2,368 most important high-frequency headwords useful for L2 learners of English, giving more than 90% coverage for most general texts in English. With regard to vocabulary knowledge relevant to academic achievement, The Academic Word List (AWL) was developed by Coxhead in 1998, covering 570 word families outside the GSL of English words (Coxhead, 2000). Hence, a learner who acquires 3,000 word families from both NGSL and the AWL will likely be equipped to understand just over 90% of English texts.

Laufer and Nation (2011) suggest that an L2 learner should have passive vocabulary knowledge of approximately 3,000 word families in the target language (TL) to engage in a daily conversation. In order for an L2 learner to understand radio interviews or literature without referring to dictionaries too often, the learner has to
possess vocabulary knowledge ranging from 6,000 to 9,000 word families. Hirsh and Nation (1992) convincingly argue that readers must be familiar with 95% of the words in a text to comprehend main points of non subject-specific writing, and in order to understand 95% of the general texts that adult learners must have knowledge of at least 5,000 word families.

A subsequent question is how many times a word must be met in the input to be learnt with its meaning. Laufer (2006) and Laufer and Nation (2011) noted that there is some chance of recognizing the meaning of a new word later only if the learner sees that word more than ten times. Besides, each new experience with the word should take place before the word is forgotten (Laufer, 2006). In a study conducted by Zahar, Cobb, and Spada (2001), the average number of words learned by their participants (i.e., ESL high school students in Quebec) while reading was 2.16 out of 30 test items. This, in turn, indicates that less than one word was learned per 1,000 words the participant read. At this rate, the authors claim that the ESL learners in the Quebec high school system should spend 29 years to learn only the most frequent 2,000 words through extensive reading. These research findings have demonstrated that vocabulary learning in an L2 is not an effortless job even for the most frequent few thousand words. It is, therefore, axiomatic that novice learners should put their first priority on the most basic and common words. Nevertheless, words that appear less frequently cannot be taken as less important for some learners. For example, those who pursue higher education in an English-speaking country need to build up higher-level vocabulary knowledge beyond the most frequent 2,000-3,000 word families, and should possess knowledge of specific vocabulary words in academic contexts as well as in general contexts. Zechmeister, Chronis, Cull, D’Anna,
and Healy (1995) indicate that the receptive size of a college-educated native English speaker is about 17,000 word families. It follows that foreign learners whose goal is to pursue higher education must have vocabulary knowledge close to that of native college graduates, and they should spend some time focusing on advanced-level, educated vocabulary.

As Coxhead (2000) confirms, the difficulty of learning academic words arises primarily due to the low-frequency of their appearance. Nation (2012, The Vocabulary Size Test) suggests three main frequency levels. The high-frequency level includes the most frequent 1,000-2,000 words, mid-frequency includes the next most frequent 3,000-9,000 words, and low-frequency includes words with a frequency level of 10,000 on.

Table 1 illustrates descriptions of high-, mid-, and low-frequency level word families (Nation 2012, p. 6).

<table>
<thead>
<tr>
<th>Level</th>
<th>1,000 word family</th>
<th>Learning procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-frequency</td>
<td>1,000-2,000</td>
<td>• Reading graded readers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliberate teaching and learning</td>
</tr>
<tr>
<td>Mid-frequency</td>
<td>3,000-9,000</td>
<td>• Reading mid-frequency readers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Deliberate learning</td>
</tr>
<tr>
<td>Low-frequency</td>
<td>10,000</td>
<td>• Wide reading</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Specialized study of a subject area</td>
</tr>
</tbody>
</table>

Following Nation’s (2012) word frequency account, most GRE words belong to low-frequency words that comprise less than 10% of the English vocabulary; however, despite their rare appearances in general contexts, they turn up more often than one
realizes in more academic contexts such as newspapers, magazines, and graduate texts. For example, recurring GRE words like *convoluted, gratuitous, or vociferous* are indeed commonly used among people in academia. Obviously, learners aiming for academic success, not to mention being admitted to a graduate school, should prioritize vocabulary acquisition in the target language so lack of vocabulary knowledge does not hamper reading comprehension.

As noted earlier, researchers have acknowledged that in order for retention to occur via incidental learning, an unknown word should be encountered at least ten times with a reasonably short interval between encounters (e.g., Brown, Waring, & Donkaewbua, 2008; Laufer, 2006; Laufer & Nation, 2011; Waring & Takaki, 2003). This still may be an achievable task in acquiring the most frequent 2,000-3,000 word families because learners encounter them often enough to recall their meaning; yet, for high-level, low-frequency words that have only 10% coverage in the English corpus, the task does not appear to be very attainable, not to mention realistic. Cobb (2007) also found that the majority of words beyond the most frequent 2,000 would not be encountered at all in a year or two even if we assume the largest plausible amounts of free reading. This being the case, it does not seem possible for learners to obtain such comprehensive vocabulary knowledge merely from extensive reading. Instead, learners are required to engage in more word-focused, deliberate and intentional learning of vocabulary words (Elgort, 2011; Groot, 2000; Laufer, 2006).

2. **L2 vocabulary instruction**

As previously emphasized, one of the most crucial aspects of language learning is
vocabulary and its pivotal role in foreign language learning has been widely recognized (e.g., Laufer & Nation, 2011; Nation, 2013; Willis & Ohashi, 2012). Virtually all L2 researchers in the field of reading ability concur that vocabulary development is an indispensable part of reading comprehension (Grabe, 1991). Nevertheless, vocabulary learning is constantly burdensome and worrisome to most L2 learners, and is more demanding than it may appear. This is largely because unlike grammar, which is made up of a restricted set of rules, vocabulary is an open set of many thousands of items. Yet, it is not just the sheer number of L2 vocabulary items to remember that makes vocabulary learning challenging. As Laufer and Nation (2011) put it, L2 vocabulary learning is difficult because of its quantitative, qualitative, and environmental/situational aspects: learners should learn features and patterns associated with vocabulary in addition to the meanings of a large quantity of words.

It is customary to think that most vocabulary items, whether in the L1 or in the L2, are acquired ‘incidentally’ as a by-product of activities in listening, reading, speaking, or writing (Hulstijn, 2001). This is learning from comprehensible meaning-focused input through listening and reading, with its main focus on understanding the information heard or read (Nation, 2013). However, it is suggested that meaning-focused learning does not lead to success in improving one’s vocabulary knowledge, and that an alternative form-focused (i.e., word-focused) learning methods should be incorporated in vocabulary teaching (e.g., Laufer, 2003, 2006). By pointing out a crucial fault in the ‘default hypothesis’ (i.e. incidental learning through reading), Laufer (2003) adds that learners do not always notice unknown words in the input. By the same token, Read (2004) states that although learners certainly acquire word knowledge incidentally while engaged in
various language learning activities, more direct and systematic study of vocabulary is also required. Nation (2013) further maintains that learning from meaning-focused input is best facilitated when the learner already knows 95% of the words in the input. This, in turn, means that learning may not occur when the learner does not know most of the vocabulary in the input. Therefore, in order for incidental learning to occur through extensive reading, vocabulary instruction should combine extensive reading with autonomous word-focused learning as suggested in Laufer and Nation (2011).

‘Incidental’ vocabulary learning is, therefore, in contrast to ‘intentional’ learning in that the former involves activities not explicitly geared to vocabulary learning while the latter involves activities aiming at committing lexical information to memory (Hulstijn, 2001). In L2 vocabulary acquisition, terms such as ‘intentional’ learning (Groot, 2000; Hulstijn, 2001), ‘deliberate’ acquisition (Lawson & Hogben, 1996), or ‘explicit’ learning (Ellis, 2008) all seem to correspond to ‘form-focused’, or ‘word-focused’ instruction. That is, intentional, though not incidental, vocabulary acquisition is commonly associated with vocabulary instruction, which will be the focus of the study described here. From a vocabulary perspective, this means that vocabulary instruction, embedded within a language course, should involve the direct teaching of vocabulary as well as the direct learning and study of vocabulary (Laufer, 2006; Nation, 2013).

Form-focused instruction (FFI) is another term that is widely used to refer to more explicit or direct teaching of language forms in isolation, as noted in the introduction to this paper. Ellis (2001) defines FFI as “any planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic form” (p.1-2), where ‘form’ stands for grammatical structures, lexical items, phonological features and
even sociolinguistic and pragmatic aspects of language. Ellis (2008) further distinguishes implicit FFI and explicit FFI, where the distinction is in the presence or absence of awareness of what is being learned. Following DeKeyser (2003 as cited in Ellis, 2008), Ellis relates explicit FFI to rules that are being thought about during the learning process. Implicit FFI, on the contrary, is meant to direct students to infer rules without noticing. Although researchers may suggest different definitions and explanations of the terminology, in the current experiment, ‘form-focused’ vocabulary instruction is referred to as ‘explicit’ vocabulary teaching in which learners are forewarned about and directed to pay attention to the linguistic forms (i.e., vocabulary items) that they are taught. A good number of studies compared meaning-focused instruction with form-focused instruction, supporting the efficacy of FFI (e.g., Elgort, 2011; Ellis, 2008; Hulstijn, 2001; Laufer, 2003; Laufer, 2006; Laufer & Nation, 2011).

Now that the benefits of form-focused teaching in L2 vocabulary acquisition have been suggested, ensuing questions arise regarding how the teacher makes learning accelerate and persist in the learner’s long-term memory. Lawson and Hogben (1996) acknowledge that use of context in L2 vocabulary acquisition is of less value, and that it is not very clear how learners associate unknown words with their meaning in reading passages. They contend that this is because writers do not necessarily make a distinction between comprehension of word meaning and acquisition of word meaning from context because the major concern of writers is to convey the meaning of the overall context (Lawson & Hogben, 1996). On this account, Lawson and Hogben addressed the importance of active, constructive elaboration of vocabulary for long-term retention: the more effectively elaborated the words are with related associations, the more likely
learners are to recall them in subsequent incidents. Correspondingly, the present study looks at how explicit teaching of vocabulary with enhanced input like multimedia cues would help promote acquisition and retention of learners’ knowledge in high-level, low-frequency English words.

3. **Multimedia learning in L2 vocabulary acquisition**

The Internet and the rise of computer-mediated communication have reshaped the uses of computers for language learning and a wide range of on-line applications are currently available for use in foreign language classes. Recent research has been looking to find out whether use of multimedia aids in a computer-mediated learning environment would enhance L2 vocabulary acquisition. Multimedia instruction is defined as “the presentation of material using both words and pictures, with the intention of promoting learning” (Mayer, 2009, p. 5). Words are presented in verbal form such as speech and printed text, and pictures are presented in pictorial form including static graphics (i.e., illustrations, graphs, photos or maps) or dynamic graphics (i.e., animation or video). Thus, multimedia applications can offer pictorial or audio-visual information in addition to traditional textual cues (Mayer, 1997).

Accordingly, multimedia instruction may refer to a wide range of methods depending on how one describes it. For instance, on-screen text, graphics or animation with sounds coming from the computer’s speaker could all be considered multimedia. Other examples of multimedia may include watching a video on a TV screen with both images and sounds, a PowerPoint presentation with graphics, or a speaker drawing

---

5 Other sources have referred to this as computer-assisted language learning (CALL) (e.g., Hubbard, 2009).
6 According to Mayer (2009), multimedia is more accurately explained as dual-mode, dual-format, dual-code, or dual-channel learning. In this study, multimedia is used to refer to all of those concepts.
something on a blackboard in a more traditional classroom setting. Even a text with printed texts and illustrations can be seen as multimedia instruction. Mayer (2009) suggests that understanding occurs when learners are able to build meaningful connections between pictorial and verbal representations. Chun and Plass (1996) stressed that retention is easier and more effective when words and phrases are presented in multiple modes. Related studies have also brought out positive results, demonstrating that visual and text together are generally more effective than either alone (Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Hubbard, 2009).

3.1. Multimedia theories

Multimedia or multimodal presentation is primarily based on theories such as the dual-coding theory (Clark & Paivio, 1991) or the cognitive load theory (Sweller, Van Merrienboer, & Paas, 1998). The dual-coding theory holds that two symbolic systems, namely the imagery system and the verbal system, mediate cognitive activity (Clark & Paivio, 1991). The imagery system processes perceptual information regarding nonverbal objects and generates mental images, while the verbal system processes linguistic information and produces speech. According to the dual-coding theory, learning can be promoted when learners utilize more than one sensory modality such as visual and verbal cues (Clark & Paivio, 1991; Mayer, 1997).

However, as human cognitive abilities are limited in capacity, the cognitive load theory (Sweller, Van Merrienboer, & Paas, 1998), which is also endorsed by the dual-coding theory, posits that visual and verbal channels are limited in capacity and that learners’ cognitive resources can become overloaded when they process more than two
sources of the same information. In other words, while mixing two modes of presentation (e.g., audio and video) maximizes the working memory capacity and increases learning, three sources of information (e.g., text, audio, and video) may cause a learner’s attention to be divided, decelerating the processing of given information. For example, when words are presented as on-screen text along with animation, the learner must process words simultaneously with animation, and this makes the visual channel compete with verbal channel. On the contrary, when words are presented as audio narration, they are processed in the verbal channel, leaving more capacity in the visual channel so that the learner can be devoted to processing the animation more deeply. This is also based on the split-attention effect (Mousavi, Low, & Sweller, 1995): when words are presented as narration, they are processed in the verbal channel, leaving more capacity in the visual channel so that the learner can be devoted to processing the animation more deeply. In this way, spoken text may reduce the load on the visual channel and increase the chances for deeper cognitive processing. As working memory can process only a few elements at one time, it is suggested that students learn more deeply when verbal and visual working memories are not overloaded (e.g., Mousavi et al., 1995).

Mayer and Moreno (2002) further postulated the cognitive theory of multimedia learning drawing on dual-coding theory and the cognitive load theory. As discussed earlier, the dual-coding theory suggests that visual and verbal information is processed in different processing systems, and the cognitive load theory suggests that the processing capacities of information are extremely limited. The cognitive theory of multimedia learning put forward some principles outlining best practices when using multimedia aids. It is suggested that using narration and animation rather than narration alone is more
effective. Words should be presented as narration rather than as on-screen text. Words and sounds should be concise, with unnecessary information eliminated. Providing on-screen text along with narration and animation is ineffective. Mayer and Moreno recommend that a learner must be actively involved in each of the cognitive processes in order for meaningful learning to occur.

In sum, multimedia theories support the idea of using multi-modal components in learning materials, yet the effective blending of components is as important as the mere presence of multimedia in learning materials. The theories advocate that learners should be provided with the proper amount of information via a limited number of modalities. Therefore, the present study will attempt to fathom ‘how and how much’ should be included in multimedia learning aids.

3.2. Multimedia annotations in computer programs

Multimedia learning materials with pictures (animation) and words (narration) offer a potential power to gear up learner understanding. However, all multimedia instructional messages are not equally effective, so the question should be how to design multimedia materials that can facilitate meaningful learning. The existing research findings with regard to multimedia aids in L2 vocabulary learning have predominantly looked into the efficacy of multimedia annotations in reading passages engrained in a computer program. Annotations or glosses, which are frequently brought up in multimedia settings, refer to a brief definition or note of a word that enhances L2 comprehension processes, and the two terms are used interchangeably in the present study. In a multimedia program, it is possible to provide a variety of annotations for words in the form of text, pictures, video,
and sound (Chun & Plass, 1996; Mohsen & Balakumar, 2011). Al-Seghayer (2001) proposed that by orchestrating various modalities such as sound, still pictures, or dynamic videos, computerized glossing has a positive effect on vocabulary acquisition as well as on reading comprehension. His study further contends that learners can immediately access different types of information without interruptions during reading, which enables them to generate causal inferences and to construct a situation model. Multimedia annotations in a computer-mediated learning environment have yielded very promising outcomes in L2 vocabulary acquisition (Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Yoshi & Flaitz, 2002).

There are studies addressing multimedia effects in more general learning contexts, rather than focusing on language acquisition. For example, Schnottz, Böckheler, and Grzondziel (1999) examined knowledge acquisition with static and animated pictures in computer-based learning. Their findings suggest that animated pictures reduce the cognitive load by providing external support for mental simulations as well as mental model construction. In contrast, static pictures support only mental model construction, but have a stronger visualization effect than animated pictures. The study suggests that pictures and text complement each other. Along similar lines, Sharp et al. (1995) reported that the helpful video condition significantly affected children’s ability to remember and retell a story to a greater degree than the minimal video condition or the no video condition. Their interpretation was that video enables children to form coherent mental models of oral stories.

In exploring the effectiveness of multimedia annotations in L2 vocabulary acquisition, Chun and Plass (1996) conducted an experiment with 103 German learners
using a hypermedia application CyberBuch for German reading texts containing multimodal annotations in the form of text, pictures, and videos. This study had a within-subjects design in which all the students used the same program. The students first watched a video that gave an overview of a German short story. Then, they individually read the story and consulted the meaning of 82 words from different types of annotations in the computer program. After reading, the students took a vocabulary test and wrote a summary of the story in their L1. The vocabulary test consisted of 36 vocabulary words and asked the students to indicate an English definition for each word: 12 words with textual definitions only (text-only), 12 with pictures and definitions (picture+text), and 12 with videos and definitions (video+text). The results demonstrated significantly higher scores for the words in the picture+text condition (31%) over the text-only condition (18%). Although the scores for the words with the video+text annotations (23%) were higher than those with text-only annotations (18%), it was not a significant difference. This study suggests the usefulness of still images presented with textual information.

A similar within-subjects experiment was conducted by Al-Seghayer (2001) to investigate the effects of different types of annotations on L2 vocabulary acquisition. Since Al-Seghayer already acknowledged the efficacy of multimodal glosses in a computer program, he specifically intended to compare static pictures with dynamic videos. Thirty intermediate ESL students with different L1 backgrounds (e.g., Arabic, Japanese, Korean, etc.) were chosen to participate in a hypermedia program. Three different modes of annotations for 15 words were used within a computer program: printed text only, printed text with still images, and printed text with dynamic videos. Each student read the story individually in the multimedia program while consulting with
the annotations for the more difficult vocabulary. After reading the story, the students took two vocabulary tests: a recognition test with multiple choice questions and a production test in which the students wrote a definition in English. Results have indicated a significant difference across the conditions, revealing the highest percentage of correct scores (87%) for words with video+text, followed by 67% for words with picture+text, and only 53% for the text alone. Even more compelling is the learners’ reaction to different types of annotations. A questionnaire and an interview administered after the treatment indicated that the video mode and the images were rated very helpful by 86.6% and 70% of participants respectively, while only 10% of the participants evaluated the text-only mode as very helpful. Al-Seghayer proposed that by orchestrating various modalities such as sound, still pictures, or dynamic videos, computerized glossing has a positive effect on vocabulary acquisition as well as on reading comprehension. His study further contends that learners can immediately access different types of information without interruptions during reading, which enables them to generate causal inferences and to construct a situation model.

By the same token, Akbulut (2007) looked into the validity of different hypermedia glosses in the reading software on vocabulary learning and reading comprehension. Sixty-nine advanced Turkish learners of English were randomly assigned to three different conditions: definition-only, definition with picture, and definition with video. A total of 42 words selected from the pilot study were embedded in three forms: text definitions, definitions with associated pictures, and definitions with videos in hypermedia reading software. Students had access to relevant information by clicking the associated links. Three types of vocabulary tests were administered: form recognition,
meaning recognition and meaning production. The form recognition test included a checklist where students indicated the words they remembered. The meaning recognition test included multiple choice questions for synonyms and definitions. The meaning production test asked the students to write the L1 or L2 equivalents or synonyms of each target word. Students took a set of three vocabulary tests immediately after the reading session and three weeks later, both unannounced beforehand. Results from the vocabulary tests indicated that the two groups who received visual annotations performed significantly better than those without visuals on both immediate and delayed post-tests. Furthermore, the video group seemed to have the best performance, the picture group ranked second, and the definition-only group had the lowest scores.

Yoshii and Flaitz (2002) investigated L2 incidental vocabulary retention in a multimedia setting with 151 adult ESL learners of beginning and intermediate levels. The participants read a short story on a computer. Fourteen target words presented with three different annotation types were provided and compared: text-only, picture-only and text and picture combined. Picture recognition, word recognition, and definition supply tests were facilitated to assess participants’ immediate and delayed retention rates. The picture recognition test asked students to choose one out of four pictures that best conveyed the meaning of each target word. In the word recognition test, students selected the definition from four choices. The definition supply test required the students to put a check-mark next to the words they remembered and to write the meaning of the word in their L1 or L2. Across the tests, the combination group outperformed the other two groups significantly in both the immediate and delayed tests, although the delayed tests manifested smaller differences than the immediate tests. The results might look
undisputable when considering differences in the amount of the input for each group: the text+picture group obviously received more input than either the text-only or the picture-only group. Yet, what is intriguing is the relative effectiveness of the picture-only annotation. It might be natural to think that picture-only group would outperform the text-only group on the picture recognition task; in the same way, the text-only group is expected to perform better on the word recognition and definition supply tests. Interestingly, however, the results showed otherwise. Though the picture-only group was not provided with textual information, they performed slightly better on the word recognition test than did the text-only group and showed a trend towards the definition supply test. Their findings exhibited no correlation between the proficiency level and the treatment type, suggesting that learners across all proficiency levels benefited from visual cues rather than conventional textual representations (Yoshii & Flaitz, 2002).

More recently, Kim and Gilman (2008) conducted an experiment to investigate the benefits of multimedia cues in a web-based self-instruction program among 172 middle school students in South Korea. In their study, the students learned 15 target words in example sentences with or without multimedia components such as text, spoken text, and graphics. They divided the students into six different groups: (1) visual text, (2) visual text and spoken text, (3) visual text and graphics, (4) visual text, graphics and spoken text, (5) reduced text and spoken text, and (6) reduced text, graphics and spoken text. The visual text includes the word’s definition with an example sentence, whereas the reduced text provides only an example sentence without the word’s definition. The spoken text is an audio recording of the target word, definition, and the example sentence. The students took a pre-test one week prior to the instruction, and an immediate post-test after the 30-
minute instruction. One week after the instruction, they were given another post-test as a retention test. The tests had 30 multiple choice questions consisting of two types: 15 questions asked for a definition of each test item and the other 15 required filling in the blank with the right word and choosing a Korean equivalent. They discovered that students in the “text with graphics” group where they had access to a definition with a graphic and the “text with audio and graphics” group in which students learned words with a definition, sound, and a graphic earned significantly higher scores than those who received other types of instruction. Their study suggested that vocabulary learning is more effective and sustainable when text was presented with graphics. From their results, Kim and Gilman concluded that graphics illustrating the meaning of words could improve learners’ knowledge of English vocabulary.

While most studies inquired into vocabulary learning and reading comprehension simultaneously when assessing the effectiveness of different types of multimedia annotations, Yanguas (2009) demonstrated contrasting influences of multimedia glosses on reading comprehension and vocabulary learning. Yanguas examined the effects of different types of multimedia glosses (i.e., textual, pictorial, and textual + pictorial) on text comprehension and vocabulary acquisition. The purpose of the study was to see whether any of the gloss conditions would facilitate noticing and whether the noticing would contribute to better comprehension. Ninety-four university students who were taking a Spanish class participated in the study and were randomly assigned to one of the four conditions: no gloss, textual gloss, pictorial gloss, or textual + pictorial gloss. A passage was extracted from an online Spanish newspaper and 21 out of 543 words were selected and glossed with a hyperlink. The participants were asked to think aloud while
reading a text under four different gloss conditions. When clicking the link, those in the
gloss groups had access to a box with a definition in English (textual gloss group), a
picture (pictorial gloss group), or a combination of the two (textual + pictorial gloss). No
gloss was provided to the control group. The assessment consisted of three tasks: a
production task, a multiple choice word recognition task, and a comprehension task.
There was a pre-test, an immediate post-test, and a delayed post-test (three weeks after
exposure) for the production and recognition tasks, and only an immediate post-test was
administered for the comprehension task. In the production test, participants were asked
to provide an equivalent (Spanish word for the given English words). In the multiple
choice recognition test, the target words in Spanish were presented with four possible
equivalents in English. The comprehension task contained 11 multiple choice questions in
English. The results showed that the participants in the gloss conditions noticed the target
words to a significantly higher degree than the control group although no significant
difference was found among any of the gloss groups. There was no significant difference
among groups in the production test of the target words. With respect to comprehension,
the combination gloss group (textual + pictorial) performed significantly better than all
other groups. The results showed the benefits of multimedia glosses, whether textual,
pictorial, or a combination of both, on recognizing new words. The results also indicated
that glosses might not offer enough support to fully learn the target words, but may
significantly help with comprehending the text. From the findings, Yanguas argues that
glosses may have different impacts on text comprehension and vocabulary learning
respectively.

The literature has mainly investigated textual and visual effects as well as the
combination of the two in computer annotations. Texts, pictures, and videos are provided so students can click on hyperlinks for further information. Although it is not clear if pictures or videos play a more crucial role in acquiring new vocabulary, the combination of textual information and visual information seems to have a more beneficial effect than either type in isolation on reading comprehension and vocabulary learning. Studies suggest that visual information stimulates textual information and that the simultaneous presentation of the two modes (i.e., visual and textual) may yield better learning outcomes.

3.3. Video effects in multimedia instruction on L2 vocabulary

Although the effect of visual information on comprehending texts in computer programs has received extensive research attention, little research has explored the impacts of visual cues on vocabulary learning in a classroom setting. One of the few such studies is Neuman and Koskinen (1992), which examined the effects of captioned television by comparing four learning conditions: captioned TV, traditional TV, reading along and listening to text, and text only. One hundred twenty-nine bilingual middle school students were measured on their degrees of word learning over a nine-week period of time. On all measures of word knowledge (i.e., word recognition, retelling story, checklist vocabulary test), the students who learned the vocabulary through viewing captioned television consistently outperformed participants in the other conditions. Moreover, the captioned television group remembered more scientific information than the rest of the participants. Their findings indicated that captioned video with sound provided a semantically enriched context, assisting participants with learning and
retaining the words better. Neuman and Koskinen also stressed that providing different kinds of information through different modalities may enhance incidental learning from context. The correlation analysis of the 90 target words further demonstrated that students learned the words most effectively when there were strong associations between a word and its video context.

Hanley, Herron, and Cole (1995) undertook a classroom study in order to compare the effects of video clips and static pictures on comprehension and retention of a written passage. Sixty-two elementary school students learning French participated in their study, and were assigned to two different conditions. The video group viewed a narration on a video, and the other group read the passage by means of pictures and the narrative of the teacher. Their findings revealed that video was a more effective method than the picture with the teacher’s narratives. They maintained that video has the potential to enhance comprehension and enrich instruction thanks to the “inherent strength of video to contextualize better than still pictures” (Hanley et al., 1995, p. 63).

Although the aforementioned research focuses on ‘visual’ effects, some studies have indicated the necessity of textual information added to visual presentation and advocated dual-mode presentation in multimedia contexts. Baltova (1999) looked at the effects of different video formats among low-proficiency learners of French: authentic video with L2 subtitles (bimodal condition), L2 subtitles with L1 audio (reversed condition), and video with no caption (traditional condition). The results showed significantly higher scores for the bimodal and reversed conditions than the traditional condition without subtitles; however, there were no significant differences between the two subtitled groups. Her study suggested that L2 learners rely more on visuals for
comprehension than native speakers do and that visuals facilitate comprehension by providing nonverbal cues through body language and actions that stimulate expectations and predictions of the message. She contends that authentic video without captioned text can provide comprehensible input, but may not be accessible for learning because of its rapid speech rate or linked pronunciation coupled with slow processing skills. Similarly, Chun and Plass (1996) stated that although video can describe the word, it cannot give a clear translation; thus, textual information is needed in addition to video so the learners can translate unknown words. Jones and Plass (2002) also stressed that multimedia should present both visual and verbal information. They compared multimedia annotations with no annotations, visual annotations, verbal annotations, and visual+verbal annotations. Their findings suggested that vocabulary was learned better with visual cues than with verbal annotations only.

Research has yielded rather incongruous results as to whether pictures or videos would better expedite L2 vocabulary acquisition. However, the overall consensus is that utilizing or combining any form of multimedia cues is more advantageous than offering mere textual information, whether it is through multimedia annotations, or in classroom instruction. Use of visual aids in language teaching is thereupon encouraged as a substitute for or in addition to textual cues (e.g., Yoshii & Flaitz, 2002). Moreover, adding textual information to video appears to promote L2 learners’ understanding of learning materials (Baltova, 1999; Chun & Plass, 1996; Jones & Plass, 2002). As seen in the literature presented so far, most studies have looked at multimedia effects on beginner and intermediate level vocabulary, paying less attention to more advanced level words. In this respect, the current study attempts to investigate the use of multimedia enhancement
as a video form in addition to textual information on L2 vocabulary instruction on high-level, low-frequency English words that are shown on the GRE.

3.4. *Linguistic cues in multimedia presentation*

Researchers have reached agreement that multimedia components have positive effects on the learning and understanding of language. However, it is possible that some multimedia components are more advantageous than others. As noted earlier in the review of studies about multimedia annotations (Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996), the fact that the simultaneous presentation of visual and textual information yielded positive results may suggest that visual information (i.e., picture or video) is effective when presented with an annotation, a brief definition or explanation of a word. This being the case, studies have addressed the usefulness of linguistic information given about an unfamiliar word in context.

Admittedly, meanings of unknown words are hard to infer in context and the contextual information may be redundant; as a result, learners may fail to connect the form of the unfamiliar word to its meaning. Some researchers (e.g., Nation, 1982; Nation & Coady, 1988) suggest that new words are learned when learners infer the meaning from the context on their own. This viewpoint is based on the assumption that learners invest more mental effort when they try to induce a solution to a problem themselves and that information acquired with more mental effort will lead to higher retention. On the contrary, other researchers (e.g., Kelly, 1989; Koster, 1985; Stip & Hulstijn, 1986 as cited in Hulstijn, 1992) argue that context rarely offers enough information for learners to infer the meaning of an unknown word successfully and that learners may learn an unknown
word incorrectly by making false inferences. Accordingly, Hulstijn (1992) raised a question as to whether letting learners infer the meaning of an unknown word or providing them with the meaning of an unknown word would better increase vocabulary learning. In this regard, he compared three inferring methods (e.g., multiple choice, concise context, and no-cue) with a meaning-given procedure (e.g., translation/synonym). Adult learners of Dutch with Turkish L1 backgrounds participated in the study. The subjects were provided with a reading comprehension task, followed by a post-test in a multiple choice format. One of the findings revealed that subjects were more likely to infer an incorrect meaning of an unknown word in L2 when no linguistic cue was available. Consequently, Hulstijn addressed the importance of elaboration on the meaning of an unknown word in incidental vocabulary learning and argued that foreign language teaching should focus on what types of cues will be most effective. These cues include L1 translation, L2 synonyms, sample sentences, multiple choice, and more.

Hulstijn, Hollander, and Greidanus (1996) also maintained that elaboration on the meaning of an unknown word may positively affect vocabulary learning. Hulstijn et al. looked into the effects of glosses or dictionary lookups in a written text and suggested that the three factors that are likely to improve incidental vocabulary learning are the presence of marginal glosses, the use of a dictionary, and frequent encounters with unknown words. Accordingly, they explored the combined influence of the reoccurrence of a word and the provision of word meaning. Seventy-eight Dutch university students who were advanced learners of French participated in the study. They read a short story in French under three different conditions: marginal glosses, dictionary use, and control. The marginal glosses group was provided with an L1 translation for the target words. The
dictionary use group was allowed to use a dictionary. The control group did not have access to either marginal glosses or a dictionary. The short story included 16 target words. After the students read the text, they answered comprehension questions. The findings suggest that having access to word meanings through glosses or dictionaries expedites incidental vocabulary learning beyond that which occurs without any linguistic information.

Chun and Plass (1996) addressed the complexity of realistic learning situations when presenting multimedia components in class. Although the primary goal of their study was to examine the effects of different types of annotations (i.e., text, pictures, and video), it is difficult to consider each of the annotation types in isolation because learning requires attention to multiple factors. For example, watching a video and taking a follow-up test would be an idealistic research design; however, realistically, as a video alone does not describe or explain the word, textual information is necessary in order for the learner to translate an unknown word.

Jones (2004) looked into the benefits of pictorial and written annotations on L2 vocabulary learning in a multimedia environment. Students were divided into four groups: a control group with no annotations, and the three treatment groups with written, pictorial, or both written and pictorial annotations while listening to text in French. Students who received any type of annotation performed better than those without any annotation on the written vocabulary recognition test and on the pictorial vocabulary recognition test. However, in most cases, students provided with written annotations, whether alone or with pictorial annotations, revealed better performance than the rest of the groups. Furthermore, it was shown that the group who had access to the written
annotations recalled more vocabulary than all the other groups on the delayed test. This may be an indication that written annotations add value to pictorial annotations, and in turn, pictorial annotations themselves may not be as effective as the combination of the two.

Plass, Chun, Mayer, and Leutner (1998) were concerned with improving learning outcomes with the aid of multimedia learning environments. One-hundred three nonnative German college students who were taking German language courses participated in the study. They read a story consisting of 762 German words in a computer program with some marked words (i.e., 82 words). Each of the marked words was provided with a text translation, and some of them had an additional picture or a video. After reading the story, students took a meaning recall test where they provided an English translation for each word. The vocabulary post-test included 24 words that were provided with both visual and verbal annotations: 12 with text and a picture and the other 12 with text and a video. The students were also asked to write a summary of the German story in English. Plass et al. discovered that learners scored higher on a written vocabulary meaning recall test when they had access to both written and pictorial annotations than when they had access to only one annotation type. Their study also revealed that written annotations had a stronger impact on vocabulary production than did pictorial annotations. In other words, if the textual information only shows a word itself without its meaning, or if the test item appears in a context where the meaning is not construed, the textual information may not be very helpful. Along those lines, in their web-based self-instruction program, Kim and Gilman (2008) also discovered that the groups who learned the target words with graphics and word definitions significantly
outperformed the rest of the groups. In their study, even though example sentences were provided in equal number to all six groups, the groups that were provided with word definitions benefited more from the learning.

The studies indicate the value of the accurate linguistic information in multimedia visual presentation and suggest that the presentation of a precise meaning of an unknown word in multimedia materials, rather than a target word in a random context, will produce a better learning outcome. In addition to looking into multimedia effects in L2 vocabulary acquisition, the present study also attempts to discover how the audio visual aids should be presented in multimedia settings; that is, what types of video (i.e., videos with visual cues or linguistic cues) would be most advantageous to learners’ vocabulary learning.
III. Pilot Study

1. Participants

The design of the experiment and test items were piloted in a preliminary study conducted during a four-week intensive GRE lecture series held in Seoul, South Korea in July 2014. Fifteen Korean students (six males and nine females; average age of 28, SD=2.93) who were preparing for graduate study in the U.S. participated in the study. Age of onset of L2 acquisition of the participants varied from 7 to 22 (M=12.60, SD=3.81)\(^7\), and their length of residence (LOR) in an English speaking country ranged from zero to ten years with its mean length of time 1.17 years (SD=2.56). The breakdown of the participant information is summarized in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Age of onset</th>
<th>LOR</th>
<th>Michigan (45)</th>
<th>Pre-test (40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Text</td>
<td>Video</td>
<td>Text</td>
<td>Video</td>
<td>Text</td>
</tr>
<tr>
<td>mean</td>
<td>28.29</td>
<td>27.88</td>
<td>13.29</td>
<td>12.00</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.87%)</td>
</tr>
<tr>
<td></td>
<td>23.91</td>
<td>2.76</td>
<td>3.95</td>
<td>3.32</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(62.22%)</td>
</tr>
<tr>
<td></td>
<td>34.14</td>
<td>33.50</td>
<td>34.14</td>
<td>33.50</td>
<td>4.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.87%)</td>
</tr>
<tr>
<td></td>
<td>1.43</td>
<td>1.00</td>
<td>1.43</td>
<td>1.00</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.58%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.00%)</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

2. Experimental design

The pilot study primarily investigated the effects of explicit teaching of vocabulary and the effects of video clips on the development of receptive knowledge of high-level, low-frequency English words that appear on the GRE in a classroom setting. To this end,

---

\(^7\) Most of the participants considered English classes in middle school at age thirteen to be their first contact with English.
the study compared two different types of instruction in a Text+Video condition and a Text-only condition by using translation (meaning recall) tasks. The pilot study addressed the following questions:

(1) Will explicit teaching of vocabulary promote L2 vocabulary acquisition of high-level, low-frequency English words among Korean learners of English?

(2) Will the use of video clips in addition to traditional textual information facilitate L2 vocabulary acquisition of high-level, low-frequency English words among Korean learners of English?

The experimental design of the pilot study is summarized in Table 3.

<table>
<thead>
<tr>
<th>Time</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>Pre-test (60) (meaning recall test)</td>
</tr>
<tr>
<td></td>
<td>→ 40 selected as final test items</td>
</tr>
<tr>
<td>Day 4</td>
<td>Definition with dictionary annotations</td>
</tr>
<tr>
<td>Day 7</td>
<td>• Instruction: Text+Video (n=8) and Text-only (n=7)</td>
</tr>
<tr>
<td></td>
<td>• Immediate post-test (post-test 1)</td>
</tr>
<tr>
<td></td>
<td>• Questionnaire + Michigan test</td>
</tr>
<tr>
<td>Day 10</td>
<td>Delayed post-test (post-test 2)</td>
</tr>
<tr>
<td>Day 19</td>
<td>Delayed post-test (post-test 3)</td>
</tr>
</tbody>
</table>

On the first day of the lecture, the entire class was given a pre-test of 60 items. Of the 60 items, 40 items were selected as final test items. The 40 test items used for the study are as follows:

---

8 All 40 final test items selected for the study were answered correctly by fewer than 10% of the participants. The pre-test scores of the participants ranged from zero to five (0-12.5%) and no significant difference in pre-test scores was found between the two groups.
On day four of the lecture series, all the participants were provided with a definition learning session where they reviewed the meaning of each test item both in English and Korean. Figure 1 below shows part of the handout to preview the definition of the test items in English and Korean.

<table>
<thead>
<tr>
<th>Word</th>
<th>IPA symbol</th>
<th>Definition (Korean)</th>
<th>Definition (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>charlatan</td>
<td>[ʃərˈlætən]</td>
<td>n. 꼼꼼이, 사기꾼</td>
<td>a person who makes elaborate, fraudulent, and often voluble claims to skill or knowledge; a quack or fraud</td>
</tr>
<tr>
<td>miff</td>
<td>[mɪf]</td>
<td>n. 발탄하기; 하찮은 다름</td>
<td>a petty quarrel or argument</td>
</tr>
<tr>
<td>repugnant</td>
<td>[rɪˈpʌɡŋt]</td>
<td>a. 싫은, 볼때한; 반감을 돋운</td>
<td>arousing disgust or aversion</td>
</tr>
<tr>
<td>beguiling</td>
<td>[bɪɡəlɪŋ]</td>
<td>a. 매혹적인, 현혹하는</td>
<td>charming or fascinating</td>
</tr>
<tr>
<td>squalid</td>
<td>[ˈskweɪld]</td>
<td>a. 더러운, 야피한; 비밀한, 젖한</td>
<td>dirty and wretched, as from poverty or lack of care</td>
</tr>
<tr>
<td>commotion</td>
<td>[kəmˈɒʃən]</td>
<td>n. 흔들다, 동요</td>
<td>a condition of turbulent motion</td>
</tr>
<tr>
<td>smug</td>
<td>[smʊɡ]</td>
<td>a. 간단 재하는, 독신적인</td>
<td>exhibiting or feeling great or offensive satisfaction with oneself or with one's situation</td>
</tr>
</tbody>
</table>

Figure 1. Definition of test items

Three days later, the fifteen participants were randomly divided into two groups: experimental group (Text+Video) and control group (Text-only) in which the students
were taught the same set of 40 test items with different teaching materials while all the other conditions remained the same (i.e., classroom, equipment, instructor, etc.). In the Text+Video condition, 40 target words were shown in video clips with subtitles. The video clips had been edited to show the gist, containing five- to twenty-second utterances or conversations. Many videos used in the study were self-explanatory. For example, some words were repeated with easier synonyms (e.g., *pulchritude* reiterated as *beauty*), and others were visually represented with gestures or facial expressions. Each video with subtitles was played one time for the Text+Video group. The same set of words with subtitles (without video) was presented in a printed material in a Text-only condition. Figure 2 shows screenshots of movie clips used for the Text+Video instruction group. Figure 3 is part of the material used for the Text-only group.

![Figure 2. Examples of video clips for the Text+Video group](image)

As you can see, I don’t look like that. That was a moment of youthful *pulchritude* that is long since past. Youthful *pulchritude*? Don’t ask me what *pulchritude* means. *Pulchritude* means beauty.

**peeve**

A: Do you really eat breakfast here every morning?
B: Pretty much. Is that a problem?
A: No, I don’t care.
B: Seems like you care. Seems like you’re peeved about it.
A: I’m not peeved.

**pulchritude**

A: As you can see, I don’t look like that. That was a moment of youthful pulchritude that is long since passed.
B: Youthful pulchritude?
A: Don’t ask me what pulchritude means.
B: Pulchritude means beauty.

Figure 3. Handout for the Text-only group

After each treatment, participants took an immediate post-test and two delayed post-tests in the form of a meaning recall task. The example of post-tests is shown in Figure 4. A questionnaire and the listening section of Michigan Test of English Language Proficiency⁹ were used to look into participants’ linguistic background and their general English proficiency.

Figure 4. Example of post-tests (meaning recall test)

---

⁹ A more detailed description of the Michigan test is given in Section V, Methodology.
3. Results

Table 4 presents the mean scores and the standard deviations of the pre- and post-tests, and the chart in Figure 5 displays the means of the tests by instruction with error bars.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean (total 40)</th>
<th>%</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text-only</td>
<td>1.429</td>
<td>(3.57%)</td>
<td>1.8127</td>
<td>7</td>
</tr>
<tr>
<td>Text+Video</td>
<td>1.000</td>
<td>(2.50%)</td>
<td>1.0690</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>1.200</td>
<td>(3.00%)</td>
<td>1.4243</td>
<td>15</td>
</tr>
<tr>
<td>Post-test1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text-only</td>
<td>13.714</td>
<td>(34.29%)</td>
<td>6.6449</td>
<td>7</td>
</tr>
<tr>
<td>Text+Video</td>
<td>20.687</td>
<td>(51.72%)</td>
<td>8.1851</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>17.433</td>
<td>(43.58%)</td>
<td>8.0863</td>
<td>15</td>
</tr>
<tr>
<td>Post-test2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text-only</td>
<td>9.500</td>
<td>(23.75%)</td>
<td>6.5320</td>
<td>7</td>
</tr>
<tr>
<td>Text+Video</td>
<td>13.500</td>
<td>(33.75%)</td>
<td>8.2851</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>11.633</td>
<td>(29.08%)</td>
<td>7.5415</td>
<td>15</td>
</tr>
<tr>
<td>Post-test3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text-only</td>
<td>13.714</td>
<td>(34.29%)</td>
<td>8.4698</td>
<td>7</td>
</tr>
<tr>
<td>Text+Video</td>
<td>19.938</td>
<td>(49.85%)</td>
<td>7.4949</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>17.033</td>
<td>(42.58%)</td>
<td>8.3162</td>
<td>15</td>
</tr>
</tbody>
</table>

Figure 5. Means of the tests by instruction with error bars
The mean scores of each group suggest that the participants in the Text+Video group outperformed those in the text-only instruction group on all the post-tests. A repeated measures ANOVA was performed in SPSS with Method of Instruction (Text+Video and Text-only) as the between-subjects factor and Time (pre-, post-, and delayed post-tests) as the within-subject factor. The results showed that there was no significant effect of method of instruction, $F(1,13) = 2.08, ns, r = .37$. Taking together the results of all tests (i.e., pre-, immediate and delayed), it does not appear that the instruction with video clips held a statistically significant advantage over the text instruction in the present experiment.

There was a significant main effect for Time, $F(3,39) = 47.49, p < .001, r = .74$. Contrasts revealed that the scores from the first post-test were significantly higher than the scores from the pre-test, $F(1,13) = 81.35, p < .001, r = .93$, the scores from the second post-test were significantly higher than the scores from the pre-test, $F(1,13) = 32.90, p < .001, r = .85$, and the scores from the third post-test were significantly higher than the scores from the pre-test, $F(1,13) = 64.41, p < .001, r = .91$. This means that if we ignore whether instruction was given by Text+Video or Text-only, then the scores the participants received on the exams were significantly affected by which post-test they were taking, and that overall, participants benefited from the instruction regardless of the method. Figure 6 shows the mean scores of the pre- and the post-tests with error bars across the participants.
The interaction effect between the tests and the instruction type approached significance, $F(3,39) = 2.38, p = .09, r = .24$, indicating that the participant scores on the tests trended towards showing an effect of method of instruction depending on the exam that they were taking. To break down this nearly significant interaction, contrasts were performed comparing performance on the tests across participants in different instruction groups. The first contrast approached significance when comparing the Text+Video group and the Text-only group to the scores of the pre-test compared to the scores of the first post-test, $F(1,13) = 4.36, p = .06, r = 50$. The second contrast revealed a non-significant difference between method of instruction and scores on the second post-test compared to scores on the pre-test, $F(1,13) = 1.53, ns, r = .32$. The third contrast looked for differences between method of instruction and scores on the final post-test compared to scores on the pre-test. This was not significant, $F(1,13) = 2.92, ns, r = .43$. The graph in Figure 7 shows the interaction effects between the tests and the method of instruction.

Figure 6. Mean scores of the pre- and post-tests with error bars
Figure 7. Interaction effects between the tests and the instruction type

Figure 7 demonstrates that participants instructed by text-only and participants instructed by Text+Video trend towards having very similar scores on the initial test, but the participants instructed by Text+Video trended towards getting higher scores on the first post-test than the participants instructed by text-only. No significant effects between the instruction type and the exams were discovered for the second post-test and the last post-test.

In addition, Pearson’s correlation coefficient was measured to examine correlation effects between the post-test scores and their Michigan test scores, and between the post-test scores and their age of onset. Results revealed no significant relationships between either of the factors and their performance on post-tests. That is, the participants’ English proficiency level and their age of onset did not serve as good indicators of performance on the post-tests. Pearson’s correlation coefficient also indicated no significant correlations between word frequency and the difficulty level of the test items. It appears
that the word frequency of each tested word did not greatly affect participants’
performance on either the pre-test or the post-tests.

4. Discussion and conclusions

The research questions of this study were: (1) whether explicit teaching of
vocabulary would promote L2 vocabulary acquisition of high-level, low-frequency
English words, and (2) whether the use of video clips in addition to the traditional textual
information would promote L2 vocabulary acquisition of high-level, low-frequency
words among Korean learners of English. In order to explore the questions, the study
compared two types of classroom instruction conditions: Text+Video and Text-only.

As for research question (1) regarding explicit teaching of vocabulary, the results
revealed that there was a significant main effect for Time: the scores for the first post-test
and the third post-test were significantly higher than the pre-test, suggesting that all
participants benefited from the instruction in general. This is in line with the body of
literature that advocates explicit teaching of L2 vocabulary (e.g., Elgort, 2011; Ellis,

Results from research question (2) regarding the use of video clips with text
indicated that performance on the tests by the participants instructed with Text+Video
was not significantly different from the improved performance of the participants
instructed by Text-only. That is, inclusion of video clips in vocabulary instruction in the
present experiment did not hold a statistically significant advantage over the text-only
instruction. The benefits of multimedia cues (e.g., Akbulut, 2007; Al-Seghayer, 2001;
Chun & Plass, 1996; Hanley, Herron, & Cole, 1995; Hubbard, 2009; Kim & Gilman,
2008; Neuman & Koskinen, 1992; Yoshii & Flaitz, 2002) were not supported in this study. However, the fact that there was a trend towards Text+Video having an effect may suggest that an improved experimental design with more subjects would affirm the efficacy of multimedia aids and the entailing theories such as the dual-coding theory (Clark & Paivio, 1991) and the generative theory of multimedia learning (Mayer, 1997).

The results may have been limited by a number of factors. First and foremost, the small group of participants in the study is not enough to perform analytic statistics. The present study had only 15 participants whereas a great deal of research conducted large-scale experiments where more than a hundred participants were involved (e.g., 129 participants in Neuman and Koskinen (1992); 133 in Montero Perez, Peters, Clarebout, and Desmet (2014); 151 in Yoshii and Flaitz (2002); 160 in Chun and Plass (1996); 172 in Kim and Gilman (2008)). As shown in the findings, the fact that the participant test scores trended towards showing an effect of method instruction may suggest that if the number of subjects were increased, the test might reach significance.

Second, it is noted that video clips provided in the study imparted different types of video support. While some videos described the target word more explicitly by giving a definition (e.g., *Pulchritude means beauty.*), others required learners to rely more on contextual information or on gestures and facial expressions (e.g., *Don’t just be blurring stuff out. I want you to really think about your answers.*). Perhaps it is more difficult for learners to process the meaning of the second type of videos more than the first type. Moreover, not all video clips may have been equally effective in conveying meanings of the new vocabulary words. For example, it could be the linguistic cues such as definitions or synonyms that helps the learner acquire the target word. Or it could be the visual cues
perceived from facial/physical expressions or backdrops that benefit the learner in remembering the target word. From this pilot study, we were not able to determine which video cue (visual, verbal, or both) had the biggest effect. Thus, the present study intends to establish which features of multimedia input will lead most successfully to L2 vocabulary acquisition by dividing the video clips into sub-categories depending on the types of cues.

Third, another factor that may have affected the results was the type of questions administered on the post-tests. L2 vocabulary acquisition is assumed to be an incremental process consisting of several levels and dimensions of knowledge, rather than all-or-nothing phenomenon (e.g., Laufer & Paribakht, 1998; Nagy & Scott, 2001). Hence, the present study included meaning recognition test in the form of multiple choice questions in addition to the meaning recall test, which was also suggested by some researchers (i.e., Montero Perez, Peters, Clarebout, & Desmet, 2014). As the main concern of the study is to investigate learners’ receptive knowledge, it may be more sensible to use a passive vocabulary knowledge test.

Finally, more information on the participants’ English proficiency, particularly vocabulary knowledge in English would have been beneficial. Since the majority of the participants did not have standardized English proficiency test scores like the TOEFL or the TOEIC, the Michigan listening test was administered as a substitute. The test may have reflected the participants’ general English proficiency level, yet it failed to evaluate their vocabulary knowledge. As a matter of fact, the Vocabulary Size Test would have been a reliable indicator of a participant’s previous vocabulary knowledge, which may have had a direct effect on learners’ progress in vocabulary learning.
5. **Implications from the pilot study**

Based on the contribution of the pilot study, several revisions were made to the present study in terms of the test items, the number of participants, classification of the types of video, and the type of test instruments. First, the final 34 test items were selected from the pilot test of 60 preliminary vocabulary words, which was administered to 24 students. To reduce the impact of learners’ prior vocabulary knowledge, only the words that were answered correctly the least often in the pilot study (less than 10% correctness) were selected. Second, as 15 participants in the pilot study did not seem to be sufficient enough to yield data that could lead to meaningful analyses, more participants (n=78) were recruited for the present study. Third, since the type of video support might have had different effects on the learning outcomes, the videos for the present study were classified based on linguistic or visual cues. To determine if a video clip contained either a linguistic cue, visual cue, or both, two native speakers of English were consulted. Fourth, in addition to the meaning recall test (i.e., L2-L1 translation) used for the pilot test, a meaning recognition test was added in the form of multiple choice questions. Finally, in order to examine the participants’ current vocabulary knowledge, the Vocabulary Size Test was conducted along with the Michigan test.
IV. Research Questions and Hypotheses

1. Research Questions

Multimedia learning materials with visual and textual information appear to offer the potential power to enhance learner understanding. However, all multimedia instructional methods may not be equally effective, so the question is how to design multimedia materials that can facilitate meaningful learning. Accordingly, the current study addresses the following questions:

(1) Will the use of multimedia presentation facilitate L2 vocabulary acquisition of high-level, low-frequency English words as proposed by the dual-coding theory? (i.e., Text-only vs. Text+Audio, Text+Video, Text+Audio+Video)

(2) What type of multimedia presentation will best facilitate L2 vocabulary acquisition of high-level, low-frequency English words? (i.e., Text+Audio, Text+Video, Text+Audio+Video)

(3) Will providing three modalities (i.e., Text+Audio+Video) result in a lower retention level than two modalities (i.e., Text+Audio, Text+Video) as predicted by the cognitive load theory?

(4) What aspect in multimedia (i.e., linguistic cues vs. visual cues) will be the most effective in L2 vocabulary acquisition and retention of high-level, low-frequency English words?\(^{10}\)

2. Hypotheses

The study investigates the effects of explicit teaching of vocabulary and the effects

\(^{10}\) Test items with a linguistic cue apply to all four groups whereas items with visual cues are relevant only to the groups provided with video (i.e., Text+Video and Text+Audio+Video).
of multimedia aids on the development of receptive knowledge of high-level, low-frequency GRE English words in a classroom setting. The study primarily looks into multimedia effects with texts, audio and video input, and the way these different types of multimedia presentation can be combined to yield better learning outcomes. The study also examines whether processing more than two simultaneous types of input is more or less effective than processing a single mode of input in L2 vocabulary learning. To this end, the study compares four different types of classroom instruction on L2 vocabulary: (1) Text-only (control group); (2) Text+Audio (no video); (3) Text+Video (no audio); and (4) Text+Audio+Video. The hypotheses of the present study are as follows:

(1) The use of multimedia presentation will facilitate L2 vocabulary acquisition of high-level, low-frequency English words as proposed by the dual-coding theory (i.e., Text-only vs. Text+Audio, Text+Video, Text+Audio+Video).

(2) The type of multimedia presentation will be significant in the facilitation of L2 vocabulary acquisition of high-level, low-frequency English words (i.e., Text+Audio, Text+Video, Text+Audio+Video).

(3) Providing three modalities (i.e., Text+Audio+Video) will result in a lower retention level than two modalities (i.e., Text+Audio) as predicted by the cognitive load theory.

(4) The type of video support (i.e., linguistic cues vs. visual cues) will be the factor deciding the most effective in L2 vocabulary acquisition and retention of high-level, low-frequency English words.
V. Methodology

1. Participants

Four four-week intensive GRE lecture series were held at a language institute located in Seoul, South Korea from July through October, 2015, from which participants were voluntarily recruited. The objective of the lecture series was to prepare students to take the GRE and eventually apply to a graduate program in the United States. The course consisted of vocabulary learning and reading comprehension practice. The four-week course aimed to teach students as many as 1,300 words within four weeks. A total of 78 students participated in the present study and received instruction under four different conditions. Participants who scored more than 10 correct (20%) on the definition pre-test (n=5) were excluded. Eventually, a final pool of 73 participants were considered for final analysis. All four groups were taught by the same instructor and studied vocabulary in the same environment in terms of the amount of time for class and the textbook.

The participants in the current study are homogeneous in terms of language background, motivation, and English learning history, and studies propose some advantages of having a uniform group of participants. First, as a great deal of academic or high-level English vocabulary is derived from Latin-related Romance languages (Nation, 2013), learning vocabulary items from the GRE would be much more challenging to Korean students who lack a background in Romance languages. Second, the type of vocabulary in question is not of interest to just any L2 English learner, but especially to learners with more specific academic goals, such as preparing for higher education in the U.S. Thus, it is essential to find a group of participants with the same academic goal and
a similar level of motivation. Furthermore, recruiting participants from outside the U.S. may have eliminated confounding factors such as age of arrival (AOA), length of residence (LOR), or formal educational history in an English-speaking country.

Several measures were taken to examine participants’ proficiency of and exposure to English. First, length of residence (LOR) was investigated because it is known to be an affecting factor in L2 acquisition. In Flege and Liu (2001), it is shown that longer LOR for late bilinguals (i.e., 3.9 - 15.5 years), combined with the input from formal education in the U.S., results in increased L2 competence. As for the participants of the current study, LOR in an English speaking country ranged from zero to nine years with its mean length of time 1.03 years (SD=0.49). In fact, 60 out of 73 participants (82%) lived in an English speaking country no more than one year. Second, the Vocabulary Size Test11 (Korean version) (Nation & Beglar, 2007) was administered in order to determine participants’ receptive knowledge of English vocabulary. Scores (n=140) ranged from 58 (41.4%) to 118 (84.2%) with the mean score 100.52 and SD 1.52. Finally, the Michigan Test of English Language Proficiency12 was adopted to determine the participants’ general proficiency level in English. The scores on Michigan test (n=45) varied from 19 to 45 (42.2-100%) with the mean score 34.49 (76.6%) and SD 1.36. The breakdown of the participant demographic information is summarized in Table 5.

---

11 See Section 3.2 for an explanation of the Vocabulary Size Test and the reason it was chosen for the study.
12 See Section 3.3 for an explanation of the Michigan Test of English Language Proficiency and the reason it was chosen for the study.
Table 5. Demographic statistics by group (n= 73)

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Fem.</th>
<th>Age (years)</th>
<th>LOR (years)</th>
<th>Vocabulary Size (n=140)</th>
<th>Michigan (n=45)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>19</td>
<td>15</td>
<td>29.79 (4.51)</td>
<td>0.66 (1.09)</td>
<td>99.58 (12.47)</td>
<td>33.16 (6.42)</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>17</td>
<td>11</td>
<td>30.00 (6.69)</td>
<td>1.63 (2.49)</td>
<td>99.94 (8.47)</td>
<td>35.71 (4.31)</td>
</tr>
<tr>
<td>Text+Video</td>
<td>18</td>
<td>8</td>
<td>28.83 (4.84)</td>
<td>0.62 (0.79)</td>
<td>99.78 (8.45)</td>
<td>33.56 (4.52)</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>19</td>
<td>18</td>
<td>26.37 (3.51)</td>
<td>1.25 (2.37)</td>
<td>102.79 (7.88)</td>
<td>35.68 (4.69)</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>52</td>
<td>28.71 (5.21)</td>
<td>1.03 (1.89)</td>
<td>100.55 (9.68)</td>
<td>34.51 (5.24)</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td>1.98 n.s.</td>
<td>1.20 n.s.</td>
<td>0.45 n.s.</td>
<td>1.26 n.s.</td>
</tr>
</tbody>
</table>

Note: The $F$ statistic is included and significance for the one-way ANOVA is indicated with an asterisk ($^{†} p<.1$  $^{*} p<.05$  $^{**} p<.01$  $^{***} p<.001$).

When comparing participants’ English proficiency level against the scores of the Michigan listening test and the Vocabulary Size Test, the majority of the participants demonstrated a high proficiency in English. The mean score of the Michigan test for the entire population was 33.51 (SD 5.24) with the highest score of 35.71 (79.4%) for the Text+Audio group and the lowest score of 33.16 (73.7%) for the Text-only group. Except for one participant who scored 19, all the other participants scored over 20.

The mean score of the Vocabulary Size Test for the whole group was 100.55 (SD 9.68), and all of the four groups scored over 99 on average, indicating that the participants have a command of over 9,900 English word families. Only two participants scored lower than the 8,000 word families (58 and 79) level. The mean score 100.55 is higher than the 8,000 word families Nation’s (2012) account states is the threshold that enables non-native speakers to cope with a wide range of unsimplified spoken and
written texts. Furthermore, Nation and Beglar (2007) suggested that non-native speaking doctoral students should possess a vocabulary of approximately a 9,000 word families. Following the interpretations of the studies, the participants for the present study can be considered proficient learners of English in terms of the number of English words they already know.

2. Materials

2.1. Test Items

In order to select test items for the current study, a total of 60 vocabulary items were tested by 24 students in a pilot study, which took place one year before the current study. This pre-test asked the participants to write the Korean equivalent for each word (i.e., L2-L1 translation) to measure their receptive English vocabulary knowledge. The 60 preliminary test items were selected from a list of GRE vocabulary compiled from words that had been on the GRE test for the last few years. It was important to select a list of words that most learners had not seen before the experiment so the learners’ previous knowledge would not influence the test scores. As Hulstijn (2003) points out, when real words are used in experimental studies, it is almost impossible to exclude some participants having at least partial knowledge of the target words. Forty items were selected for the pilot study and 34 items were used for the present study. To minimize the effect of learners’ prior knowledge, the words that students correctly defined the least were selected with the cutoff point of 10% correctness. The 34 target words are as follows:

beguiling, coax, commotion, convivial, exasperation, flinch, fluster, garbled, gloat,
gobble, gossamer, grovel, hubris, lugubrious, miffed, nonchalant, pamper, perturbed, petulant, pristine, pulchritude, repugnant, smug, solder, squabble, squalid, staunchly, sty mie, swerve, torment, toupee, vitiate, wanton, yank

None of the 34 test items were included in the most frequent 5,000 words in the Corpus of Contemporary American English (COCA) suggested by www.wordfrequency.info (Davies, n.d.). In fact, the frequency of each word was further checked off using the word rank proposed by www.wordcount.org (Harris, 2003), which covers 86,800 words based on the British National Corpus. The word frequency for the 34 test items ranged from 14,641 for torment to 84,353 for fluster (mean=36,021; SD=15,404) except for pulchritude which is beyond the 86,800 word list (see Appendix A for full description). Figure 8 shows the word frequency of the 34 test items.

---

Figure 8. Word frequency for 34 test items

---

13 Word frequency here refers to the rank based on the number of occurrences of a given word in the corpus. For example, word frequency 150 means the 150th most frequently used word. Thus, the higher the number is, the less frequent the word is used.
2.2. **Video clips**

A total of 34 video clips extracted from movies or TV shows were used for the multimedia presentation of the 34 test items (see Appendix A for subtitles). To depict meaning precisely, each video clip was edited to show the gist of the target word, containing five- to twenty-second utterances or conversations. Further, in order to examine learning effects of different types of video, the video clips were sub-divided into different categories according to whether they contained linguistic cues and/or visual cues to the meaning of the target word. For example, linguistic cues consist of definitions (e.g., *pulchritude* defined as *beauty*) and synonyms (e.g., *beguiling* reiterated as *lovely*). Visual cues include physical gestures, facial expressions, or backdrops that depict the meaning of the word. Two native-speakers of English were consulted to identify whether the video offers a ‘linguistic’ cue or a ‘visual’ cue. To do this, the native speakers first read the subtitles without watching the video to determine whether there are any linguistic cues that would lead to comprehension of the word. Next, the video clips were shown without audio to determine whether any visual clue alone would lead to the meaning of the word. The 34 video clips were subdivided into four groups accordingly: (1) both linguistic and visual cues (LVC, n=8) (+ling, +visual); (2) linguistic cue without visual cue (LC, n=8) (+ling, -visual); (3) visual cue without linguistic cue (VC, n=9) (-ling, +visual); (4) neither linguistic nor visual cues (NC, n=9) (-ling, -visual) (See Appendix A for types of video).
3. Instruments

3.1. Pre-test and post-tests

A pre-test was given before instruction to measure the participant’s current vocabulary knowledge of the target words. In order to examine the effect of instruction, two post-tests were administered after each treatment: one immediately subsequent to instruction and another approximately seven days after each treatment. The pre- and post-tests were twofold: a meaning recall test (i.e., L2-L1 translation) and a meaning recognition test (i.e., multiple choice questions). As a type of test that is adopted by studies looking at reading comprehension ability (e.g., Akbulut, 2007; Al-Seghayer, 2001; Yoshi & Flaitz, 2002), the meaning recall test asks students to write a definition of each word either in their L1 or L2. The meaning recognition test is in the form of multiple choice questions in English as suggested in a number of studies (e.g., Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Yoshi & Flaitz, 2002). Following Read and Chapelle’s (2001) framework, the meaning recognition test is discrete, selective and context-independent, and is presented in a multiple choice format. That is, the target words are presented in isolation without context. This type of test is to determine whether the learner really knows the words without having a chance to infer the meaning from contextual cues, and is suitable to measure receptive knowledge of vocabulary and to assess the effectiveness of systematic vocabulary learning (Read, 2000). In addition to the 34 test items, six distractors\(^\text{14}\) were included to make a total of 40 items in the pre- and post-tests. The sequence of each of these tests was randomized. Students took the meaning recall test before the meaning recognition test for all pre- and post-tests so the

\(^{14}\) The six distractors are banter, besmirch, fling, gripe, strut and vindictive.
answer choices in the meaning recognition test would not affect their decision making (See Appendix C and D for the full-length tests).

3.2. *Vocabulary size test*

In order to determine participants’ current knowledge of English vocabulary and to correlate with their post-test scores, Nation and Beglar’s (2007) The Vocabulary Size Test (Korean version) was adopted. The Vocabulary Size Test is designed to measure a learner’s written receptive vocabulary size in English. The test consists of 140 multiple choice items, with ten words from each 1,000 word family level. A learner’s receptive vocabulary size can be calculated by multiplying the score by 100. Each tested word appears in a single non-defining context, requiring relatively developed knowledge about the tested words.

Studies using the Vocabulary Size Test reveal that non-native English speakers need to have a vocabulary around 5,000-6,000 word families in order to be able to study at an undergraduate program at an English speaking university. Non-native speaking doctoral students possess approximately a 9,000 word family vocabulary (Nation & Beglar, 2007).

As recommended by Nation and Beglar (2007), 40 minutes was given for the participants to answer 140 items, allowing them to have an ample amount of time to ponder over each item. The bilingual version (i.e., Korean version) of the test was used for the current study because it is suggested that translations avoid the difficult grammar of English definitions, enabling leaners to comprehend them more immediately (Nation & Beglar, 2007). Figure 9 below shows examples of the Vocabulary Size Test (Korean version).
Figure 9. Examples of the Vocabulary Size Test (Korean version)

### Vocabulary Size Test (Korean version)

**First 1000**

1. see: They saw it.
   a. 잘못다
   b. 기다렸다
   c. 보았다
   d. 시작했다

2. time: They have a lot of time.
   a. 돈
   b. 음식
   c. 시간
   d. 친구들

3. period: It was a difficult period.
   a. 질문
   b. 기간
   c. 해야 할 일
   d. 책

3.3. *Michigan Test of English Language Proficiency*

In order to evaluate participants’ general English proficiency, Michigan Test of English Language Proficiency was administered. In this listening comprehension test, participants listened to a total of 45 problems, either questions or statements, and selected the best possible answer out of three answer choices (See Appendix F for the full-length test). The length of the test was approximately 20 minutes.

3.4. *Questionnaire*

After the treatment, participants were asked to complete a questionnaire, which includes questions regarding participants’ language learning background such as age of onset, length of studying English, length of residence (LOR) and education in an English-
speaking country, self-evaluation of English proficiency. They also rated the instruction based on the quality and the usefulness. (See Appendix E).

4. Treatments

All the groups studied the definitions of each word in both English and Korean using dictionary annotations (see Appendix B) before their treatment. The definitions were presented on the projector with the instructor reading each target word and its definition. Google Dictionary and Kum Sung’s New Ace English-Korean Dictionary (1980) were used as the sources of the definitions. Figure 10 shows the definitions for the first five target words.

<table>
<thead>
<tr>
<th>Word</th>
<th>IPA symbol</th>
<th>Definition (Korean)</th>
<th>Definition (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>beguiling</td>
<td>/bɪɡˈaɪlɪŋ/</td>
<td>adj. 매혹적인, 현혹하는</td>
<td>charming or enchanting, sometime in a deceptive way</td>
</tr>
<tr>
<td>coax</td>
<td>/ˈkoʊks/</td>
<td>v. (남을) 말해서 -하도록 하다, 부추기다</td>
<td>to persuade or try to persuade by pleading or flattery</td>
</tr>
<tr>
<td>commotion</td>
<td>/kəmˈɒʃən/</td>
<td>n. 동요, 황분: 소란</td>
<td>a state of confused and noisy disturbance</td>
</tr>
<tr>
<td>convivial</td>
<td>/ˌkɒnvɪˈvɪl/</td>
<td>adj. 친근한, 명랑한, 우호적인</td>
<td>cheerful and friendly; sociable</td>
</tr>
<tr>
<td>exasperation</td>
<td>/ˌɛksəˈspɛrəˈeɪʃən/</td>
<td>n. 분개, 격노</td>
<td>the feeling of intense irritation or annoyance</td>
</tr>
</tbody>
</table>

Figure 10. Screenshots of definitions of target words

For all four conditions, subtitles for each video clip were provided whether or not they accompanied video or audio. As discussed earlier, Baltova (1999) maintains that
authentic video without caption may not be accessible for learning because of its rapid speech rate or linked pronunciation. This being the case, the learners may have difficulty understanding authentic video without prior treatment because the selected target words are mostly unfamiliar to learners. Hence, this learning process is expected to help students better understand the video clips without their listening ability affecting the test results.

The four conditions are as follows:

(1) ‘Text-only’ condition: after previewing a set of five vocabulary words, the participants read five corresponding subtitles presented on the projector screen. Each subtitle was shown on a single separate slide with the target word underlined.

(2) ‘Text+Audio’ condition: after previewing a set of five vocabulary words, the participants read five corresponding subtitles presented on the projector screen while listening to the corresponding audio (without the video). Each subtitle was shown on a single separate slide with the target word underlined.

(3) ‘Text+Video’ condition: after previewing a set of five vocabulary words, the participants read five corresponding subtitles presented on the projector screen while watching the corresponding video (without the audio). Each subtitle and the video were shown on a single separate slide with the target word underlined.

(4) ‘Text+Audio+Video’ condition: after previewing a set of five vocabulary words, the participants read five corresponding subtitles presented on the projector screen while watching the corresponding video with audio. Each subtitle and video was shown on a single separate slide with the target word underlined.
Figure 11 represents how the target word *beguiling* was presented in all four conditions.

![Screenshots of target word in subtitles](image)

5. Procedures

The experiment consisted of a pre-test, a treatment with four different conditions, and two post-tests. On the first day, a vocabulary size test and an English proficiency test were administered prior to the pre-test to look into participants’ general knowledge in English. Then, participants were given two pre-tests, one definition test and one multiple choice test, for 40 items (34 test items and 6 distractors) in order to gauge the vocabulary knowledge of the test items before instruction. Only the students who did not score at
ceiling (over 20%) were selected for further tests. Then, the participants were randomly divided into four groups: one control (Text-only) and three experimental groups (Text+Audio, Text+Video, Text+Audio+Video) in which the students studied the same set of 34 test items. In each treatment, definitions were presented on the screen for every five words before instruction. A slide containing definitions for five target words was presented, followed by instruction with Text-only, Text+Audio, Text+Video or Text+Audio+Video on the same five words. One immediate post-test (post-test 1) was administered right after each treatment, and a delayed post-test (post-test 2) was given seven days after the treatment. In addition, the participants filled out the questionnaire about their language learning background and attitudes towards the given instruction. All other conditions such as classroom, equipment, and instructor remained the same for each group. Table 6 shows the experimental design of the study.

Table 6. Experimental design

<table>
<thead>
<tr>
<th>Time</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>• Vocabulary size test + Michigan test</td>
</tr>
<tr>
<td></td>
<td>• Pre-tests on 40 items (34 test items + 6 distractors)</td>
</tr>
<tr>
<td>Day 2</td>
<td>• Definition preview</td>
</tr>
<tr>
<td></td>
<td>• Treatments (instruction):</td>
</tr>
<tr>
<td></td>
<td>(Text-only; Text+Audio; Text+Video; Text+Audio+Video)</td>
</tr>
<tr>
<td></td>
<td>• Immediate post-tests (post-test 1)</td>
</tr>
<tr>
<td>Day 9</td>
<td>• Delayed post-tests (post-test 2)</td>
</tr>
<tr>
<td></td>
<td>• Questionnaire</td>
</tr>
</tbody>
</table>
VI. Analyses

1. Participant analysis

Scores on the Vocabulary Size Test and the Michigan Test of English Language Proficiency along with demographic statistics, number of participants, gender, age, length of residence (LOR) in an English-speaking country were presented in Table 5 in Section V, Methodology.

Since the participants were randomly assigned to the conditions, the groups were considered to be equal in all aspects. However, a further comparison of the groups on age, length of residence (LOR) in an English-speaking country, vocabulary size, and English proficiency level was conducted in order to determine whether the groups differed from each other in terms of these confounding variables. Additionally, scores on the pre-tests were also compared across groups. One-way ANOVAs were run to explore the instructional group differences.

1.1. Age

A one-way ANOVA was run to see if there was a significant difference in age between instruction groups. There was a non-significant main effect of group, F(3, 69)=1.98, p=.12, $\eta^2=.08$. No post-hoc t-tests were run as the main effect of group was not significant.

1.2. Length of residence (LOR)

A one-way ANOVA was run to see if there was a significant difference in length of residence (LOR) in an English-speaking country between instruction groups. There was a
non-significant main effect of group, $F(3,69)=1.20$, $p=.32$, $\eta^2=.05$. No post-hoc t-tests were run as the main effect of group was not significant.

1.3. **Vocabulary size test**

A one-way ANOVA was run to see if there was a significant difference in vocabulary size between instruction groups. There was a non-significant main effect of group, $F(3,69)=0.45$, $p=.72$, $\eta^2=.02$. No post-hoc t-tests were run as the main effect of group was not significant.

1.4. **Michigan Test of English Language Proficiency**

A one-way ANOVA was run to see if there was a significant difference in test score on the Michigan test of English Language proficiency between instruction groups. There was a non-significant main effect of group, $F(3,69)=1.25$, $p=.30$, $\eta^2=.05$. No post-hoc t-tests were run as the main effect of group was not significant.

1.5. **Definition pre-test**

Mean score and percent (score out of 34) by group for the definition pre-test is seen in Table 7.
Table 7. Mean score and percent by group for definition pre-test (n=34)

<table>
<thead>
<tr>
<th>Group</th>
<th>Score (SD)</th>
<th>Percent (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>1.32 (2.24)</td>
<td>3.87 (6.58)</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>1.71 (2.37)</td>
<td>5.02 (6.96)</td>
</tr>
<tr>
<td>Text+Video</td>
<td>1.67 (2.74)</td>
<td>4.90 (8.07)</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>2.42 (2.32)</td>
<td>7.12 (6.81)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1.78 (2.41)</td>
<td>5.24 (7.07)</td>
</tr>
</tbody>
</table>

Note: The F statistic is included and significance for the one-way ANOVA is indicated with an asterisk (*p<.05  **p<.01  ***p<.001).

To explore group differences for percent score on the definition pre-test, a one-way between-subjects ANOVA was run and there was a non-significant main effect of group, F(3,69)=0.70, p=.558, \( \eta^2 = .03 \). None of the groups scored significantly higher than the other groups, and each group showed the same level of performance on the definition pre-test.

To further investigate group differences in terms of LOR, Michigan score, and Vocabulary size score, Pearson’s Correlations were conducted. A table of correlations between participant score and length of residence, vocabulary size, and Michigan test score by group can be seen in Table 8.
Table 8. Pearson’s Correlations between percent score and LOR, Michigan score, and vocabulary size on definition pre-test

<table>
<thead>
<tr>
<th>Group</th>
<th>LOR</th>
<th>Michigan Score</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>.52</td>
<td>.22</td>
<td>.21</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>.26</td>
<td>.36</td>
<td>.40</td>
</tr>
<tr>
<td>Text+Video</td>
<td>-.03</td>
<td>.002</td>
<td>.58*</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>-.10</td>
<td>.36</td>
<td>.58**</td>
</tr>
</tbody>
</table>

† p<.10, * p<.05, ** p<.01, *** p<.001

A scatterplot of length of residence by percent score by group with linear regression line can be seen in Figure 12.

Figure 12. Scatterplot of participants’ percent score and length of residence by group with linear regression line

A scatterplot of Michigan score by percent score by group with linear regression line can be seen in Figure 13.
A scatterplot of vocabulary size by percent score by group with linear regression line can be seen in Figure 14.

1.6. *Multiple choice pre-test*

Mean score and percent (score out of 34) by group for the multiple choice pre-test is seen in Table 9.
Table 9. Mean score and percent by group for multiple choice pre-test (n=34)

<table>
<thead>
<tr>
<th>Group</th>
<th>Score (SD)</th>
<th>Percent (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-Test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text-only</td>
<td>4.95 (4.95)</td>
<td>14.55&lt;sup&gt;a&lt;/sup&gt; (14.56)</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>9.59 (3.73)</td>
<td>28.20&lt;sup&gt;a,b&lt;/sup&gt; (10.96)</td>
</tr>
<tr>
<td>Text+Video</td>
<td>5.22 (3.84)</td>
<td>15.36&lt;sup&gt;b&lt;/sup&gt; (11.30)</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>6.63 (5.71)</td>
<td>19.50 (16.79)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>6.53 (4.91)</td>
<td>19.22 (14.45)</td>
</tr>
<tr>
<td>F</td>
<td>3.64&lt;sup&gt;*&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

Note: Within the predictor on both dependent variables, two categories share a common superscript if their difference is statistically significant at either .05, .01, or .001 level (“a” or “b” indicate p<.05, “aa” or “bb” indicate p<.01, and “aaa” or “bbb” represent p<.001). Those compared means without a common superscript do not differ from each other at any of the levels of statistical significance considered.

Note: The F statistic is included and significance for the one-way ANOVA is indicated with an asterisk (↑ p<.1     * p<.05     ** p<.01     ***p<.001).

To explore group differences for percent score on the multiple choice pre-test, a one-way between-subjects ANOVA was run and there was a significant main effect of group, F(3,69)=3.64, p=.017, η²=.13. Pairwise t-tests with a Holm-Bonferroni correction were run. The Text+Audio group (M=28.20, SD=10.96) had a significantly higher percentage (all p’s<.05) on the multiple choice pre-test than the Text-only group (M=14.55, SD=14.56) and the Text+Video group (M=15.36, SD=11.30). The percent score on the pre-test of the Text+Audio+Video group (M=19.50, SD=16.79) was not significantly different (all p’s>.05) from the Text-only group, Text+Audio group, and the Text+Video group. The Text+Video group was not significantly different (p>.05) from the Text-only group.
To further investigate group differences in terms of LOR, Michigan score, and vocabulary size score, Pearson’s Correlations were conducted. A table of correlations between participant score and LOR, vocabulary size, and Michigan score by group can be seen in Table 10.

Table 10. Pearson’s Correlations between percent score and LOR, Michigan score, and vocabulary size on multiple choice pre-test

<table>
<thead>
<tr>
<th>Group</th>
<th>LOR</th>
<th>Michigan Score</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>.35</td>
<td>.16</td>
<td>.33</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>.10</td>
<td>.14</td>
<td>.38</td>
</tr>
<tr>
<td>Text+Video</td>
<td>.31</td>
<td>.01</td>
<td>.44†</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>.42†</td>
<td>.52*</td>
<td>.53*</td>
</tr>
</tbody>
</table>

† p<.10, * p<.05, ** p<.01, *** p<.001

A scatterplot of length of residence by percent score by group with linear regression line can be seen in Figure 15.

![Figure 15. Scatterplot of participants’ percent score and LOR by group with linear regression line](image)

A scatterplot of Michigan test score by percent score by group with linear regression line can be seen in Figure 16.
Figure 16. Scatterplot of participants’ percent score and Michigan score by group with linear regression line

A scatterplot of vocabulary size by percent score by group with linear regression line can be seen in Figure 17.

Figure 17. Scatterplot of participants’ percent score and vocabulary size by group with linear regression line

Participant analysis was conducted to learn if participants were distributed evenly in each group. The results of the analyses revealed that the groups did not differ from each other on age, length of residence (LOR) in an English-speaking country, vocabulary size, and Michigan score. There was no significant main effect of group on the definition
pre-test scores, yet there was a significant effect of group on the multiple choice pre-test scores. The Text+Audio group performed significantly better than the Text-only group and the Text+Video group. However, this is not expected to affect the data analysis because the study looks at the percent differences between the pre-test and the post-tests.

2. Performance analysis

As already mentioned, two different types of tests (definition and multiple choice) were given at three different times. The performance on each test by group is presented in Section VII, Results as the raw score and percent (score out of 34). The main dependent measure of interest is percent difference. The percent difference is the difference between percent on the first post-test and the pre-test, and the second post-test and the pre-test. The mean scores, percent, and percent differences are plotted and presented by test in the following sections.

2.1. ANOVAs

A one-way between-subject factorial ANOVA was run to look at significant differences between instructional groups on percent difference on all target items. One-way ANOVAs were also run to see if there were significant differences in participant information (i.e., age, LOR, vocabulary size test, and the Michigan) and in post-experiment rating of instruction by group.
2.2. *Logistic mixed effects modeling*

Participant performance, for the definition and multiple choice tests, was modeled using logistic mixed-effects models with random intercepts for item and participant. The dependent variable used for modeling was *improvement* (answering incorrectly on the pre-test and correctly on the post-test) versus *no change* (answering correctly on the pre-test and correctly on the post-test, or answering incorrectly on the pre-test and incorrectly on the post-test) or *worse performance* (answering correctly on the pre-test and incorrectly on the post-test).

There were two types of variables included in the modeling: experimental variables of interest that are the focus of the research questions and variables that were included in the model to control for any non-hypothesis related variance. The main experimental variables of interest included in the model were: audio instruction, video instruction, and the interaction between audio and video instruction. Additional experimental variables of interest included in the model testing the hypothesis that cue-type has a significant effect on vocabulary acquisition were: linguistic cue, visual cue, and the interaction between linguistic and visual cue.

Although not queried in the research questions of the study, item-level variables and subject-level variables were considered in the analysis in case these factors have affected participant performance. To explore the effect of the items themselves, variables looking at the length of the target words in terms of letters and part of speech of the vocabulary items were included. To explore the effect of prior exposure, experience, and proficiency in English, variables of length of residence in an English-speaking country, vocabulary size, and score on Michigan test were included in the model. Demographic
variables, gender and age, were included to control for any confounds unrelated to the research questions. Random intercepts (but not slopes) were included for participant and item to model individual participant and individual item variation.

The model of participant improvement that is reported for each post-test contains all of the variables so that the statistical significance of each experimental variable of interest can be examined while controlling for non-hypothesis related variance. The fit of the model was compared to the null model and the chi-squared statistic is reported.

2.3. Correlations

Correlations between participant percent difference by test, by group and length of residence in an English-speaking country, Michigan score, and vocabulary size were examined and are presented below as scatter plots. Pearson’s correlation coefficients are also presented by group in a table in each section.

3. Item analysis

Performance on the pre-test across type (definition and multiple choice) by target item was examined to establish that no one item was easier than the others. The items were similarly difficult, M=3.00, SD=3.68, range: 0-13.33. Item difficulty was calculated by looking at mean performance by item across the post-tests. Item difficulty followed a normal distribution, M=38.03, SD=18.27, range: 11.11-77.78.
VII. Results

To test the unique effect of audio instruction and the unique effect of video instruction, we decomposed the four groups using two conditions: [± Audio Instruction] and [± Video Instruction]. This is illustrated in Table 11 below. These two conditions and their interaction are the experimental variables used in the analytic statistics reported in the further sections.

Table 11. 2 x 2 matrix of instructional conditions and the 4 experimental groups

<table>
<thead>
<tr>
<th>Video Instruction</th>
<th>Audio Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-]</td>
<td>[-]</td>
</tr>
<tr>
<td>Text-only</td>
<td>T</td>
</tr>
<tr>
<td>[+]</td>
<td>[+]</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>TA</td>
</tr>
<tr>
<td>[+]</td>
<td></td>
</tr>
<tr>
<td>Text+Video</td>
<td>TV</td>
</tr>
<tr>
<td>[+]</td>
<td></td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>TAV</td>
</tr>
</tbody>
</table>

1. Definition test analysis

1.1 Definition post-test 1

Mean score, percent (score out of 34), and percent change between pre-test and post-test 1 by group for the definition test is seen in Table 12. Mean percent change by group is presented with 95% confidence interval standard error bars in Figure 18.
As Table 12 above shows, multimedia groups in general had higher mean percent change than the Text-only group (34.37%), with the highest Text+Audio+Video (45.20%) followed by Text+Audio (44.81%) and Text+Video (43.46%). In order to determine whether these differences were statistically significant, a one-way between-subjects
ANOVA was run and there was a non-significant main effect of group, $F(3,69)=2.44$, $p=.07$, $\eta^2=.10$. In the definition post-test 1, no one group outperformed the rest of the groups.

A table of correlations between participant score and LOR, vocabulary size, and Michigan test score by group can be seen in Table 13.

Table 13. Pearson’s Correlations between percent change and LOR, Michigan score, and vocabulary size on definition post-test 1

<table>
<thead>
<tr>
<th>Group</th>
<th>LOR</th>
<th>Michigan Score</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>.22</td>
<td>.55*</td>
<td>.62**</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>-.45†</td>
<td>-.21</td>
<td>.02</td>
</tr>
<tr>
<td>Text+Video</td>
<td>.35</td>
<td>.54*</td>
<td>.39</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>-.09</td>
<td>.15</td>
<td>-.17</td>
</tr>
</tbody>
</table>

† $p<.10$, * $p<.05$, ** $p<.01$, *** $p<.001$

A scatterplot of length of residence by percent difference between the definition pre-test and post-test 1 and length of residence by group with linear regression line can be seen in Figure 19.

Figure 19. Scatterplot of participants’ percent difference between the definition pre-test and post-test 1 and LOR by group with linear regression line
A scatterplot of Michigan score by percent difference between the definition pre-test and post-test 1 and Michigan test score by group with linear regression line can be seen in Figure 20.

![Figure 20](image)

Figure 20. Scatterplot of participants’ percent difference between the definition pre-test and post-test 1 and Michigan score by group with linear regression line

A scatterplot of vocabulary size by percent difference between the definition pre-test and post-test 1 by group and vocabulary size test with linear regression line can be seen in Figure 21.

![Figure 21](image)

Figure 21. Scatterplot of participants’ percent difference between the definition pre-test and post-test 1 and vocabulary size by group with linear regression line
The analysis considered experimental variables of interest (i.e., instruction type and cue type) and other variables to control for any non-hypothesis related variance (i.e., item variable and subject variable). The logistic mixed effects model with random intercepts for item and participant of participant improvement on the first definition post-test is presented below in Table 14. The model containing 13 predictor variables\(^{15}\) and two 2-way interactions\(^{16}\) was a significantly better fit than the base model without any predictor variables, \(\chi^2(16)=42.06, p<.001\).

\(^{15}\) The 13 predictor variables are: Instruction.Audio, Instruction.Video, Cue.Ling(ing), Cue.Vis(ual), Word.Frequency, Word.Length, PartOfSpeech, NumberOfWords, Gender, Age, LOR, Vocab, and Michigan.

\(^{16}\) The two-way variables are: Instruction(Audio):Instruction(Video) and Cue.Ling(ing):Cue.Vis(ual)
Table 14. Logistic mixed-effects model of participant improvement on definition post-test 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.52</td>
<td>0.37</td>
<td>-1.40</td>
<td>.163</td>
</tr>
<tr>
<td><strong>Experimental Variables: Instruction Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruction.Audio</td>
<td>0.51</td>
<td>0.22</td>
<td>2.33</td>
<td>.020 *</td>
</tr>
<tr>
<td>Instruction.Video</td>
<td>0.49</td>
<td>0.22</td>
<td>2.21</td>
<td>.027 *</td>
</tr>
<tr>
<td>Instruction.Audio:Instruction.Video</td>
<td>-0.49</td>
<td>0.31</td>
<td>-1.58</td>
<td>.114</td>
</tr>
<tr>
<td><strong>Experimental Variables: Cue Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cue.Ling(ling)</td>
<td>0.68</td>
<td>0.32</td>
<td>2.09</td>
<td>.037 *</td>
</tr>
<tr>
<td>Cue.Vis(visual)</td>
<td>0.16</td>
<td>0.32</td>
<td>0.50</td>
<td>.616</td>
</tr>
<tr>
<td>Cue.Ling(ling):Cue.Vis(visual)</td>
<td>-0.46</td>
<td>0.44</td>
<td>-1.06</td>
<td>.289</td>
</tr>
<tr>
<td><strong>Item-level Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word.Frequency</td>
<td>-0.17</td>
<td>0.12</td>
<td>-1.36</td>
<td>.175</td>
</tr>
<tr>
<td>Word.Length</td>
<td>0.18</td>
<td>0.17</td>
<td>1.02</td>
<td>.307</td>
</tr>
<tr>
<td>PartOfSpeech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POS(verb)</td>
<td>-0.25</td>
<td>0.39</td>
<td>-0.64</td>
<td>.525</td>
</tr>
<tr>
<td>POS(adj)</td>
<td>-0.55</td>
<td>0.32</td>
<td>-1.74</td>
<td>.082 †</td>
</tr>
<tr>
<td>NumberOfWords(subtitle)</td>
<td>-0.21</td>
<td>0.19</td>
<td>-1.13</td>
<td>.259</td>
</tr>
<tr>
<td><strong>Subject-level Variables: Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender(M)</td>
<td>-0.25</td>
<td>0.17</td>
<td>-1.50</td>
<td>.134</td>
</tr>
<tr>
<td>Age</td>
<td>0.00</td>
<td>0.08</td>
<td>-0.01</td>
<td>.996</td>
</tr>
<tr>
<td>LOR</td>
<td>-0.07</td>
<td>0.08</td>
<td>-0.86</td>
<td>.388</td>
</tr>
<tr>
<td>Vocab</td>
<td>0.26</td>
<td>0.09</td>
<td>2.82</td>
<td>.005 **</td>
</tr>
<tr>
<td>Michigan</td>
<td>0.19</td>
<td>0.10</td>
<td>1.94</td>
<td>.052 †</td>
</tr>
</tbody>
</table>

† p<.10,  * p<.05,  ** p<.01,  *** p<.001

**Formula in R:** Answer ~ Instruction.Audio + Instruction.Video + Instruction.Audio:Instruction.Video + Frequency + Cue.Ling + Cue.Vis + Cue.Ling:Cue.Vis + Length + POS + NumberOfWords + Gender + Age + LOR + Vocab + Michigan + (1 | Participant) + (1 | Item)

For the instructional variables of interest, participants who received audio instruction, the Text+Audio and Text+Audio+Video groups, were significantly more likely to improve on the first definition post-test on average than participants who did not receive audio instruction, the Text-only and Text+Video groups, $\beta=0.51$, SE($\beta$)=0.49,
z=2.33, p<.05. Participants who received video instruction, the Text+Video and Text+Audio+Video groups, were significantly more likely to improve on the first definition post-test on average than participants who did not receive video instruction, the Text-only and Text+Audio groups, \( \beta=0.49, \text{SE}(\beta)=0.22, z=2.21, p<.05 \). The two-way interaction of audio instruction and video instruction was not significant, \( \beta=-0.49, \text{SE}(\beta)=0.31, z=-1.58, p=.11 \).

For the cue variables of interest, participants were significantly more likely to improve on linguistic cue items than items without linguistic cues, \( \beta=0.68, \text{SE}(\beta)=0.32, z=2.09, p<.05 \). Participants were not significantly more likely to improve on items which contained visual cues than items without visual cues, \( \beta=0.16, \text{SE}(\beta)=0.32, z=0.50, p=.62 \). The two-way interaction of linguistic cue and visual cue was not significant, \( \beta=-0.46, \text{SE}(\beta)=0.44, z=-1.06, p=.29 \).

For the item-level variables, none of the variables significantly predicted whether a participant would improve on an item in the first definition post-test and only part of speech (POS) of the item approached being a significant predictor of participant improvement. Word frequency in English, word length (i.e., the number of letters in the target word), and number of words in the subtitle were not significant predictors in modeling participant improvement on definition post-test 1. Participants trended being more likely to improve on nominal items than adjectival items (\( \beta=-0.55, \text{SE}(\beta)=0.32, z=-1.74, p=.08 \)) but were not significantly less likely to improve on nominal items than verbal items (\( \beta=-0.25, \text{SE}(\beta)=0.39, z=-0.64, p=.53 \)).

For the subject-level, demographic variables, only vocabulary size was a significant predictor in modeling participant improvement. The larger vocabulary size a participant
had, the more likely they were to improve on the post-test, $\beta=0.26$, $SE(\beta)=0.09$, $z=2.82$, $p<.01$. Participant score on the Michigan approached significance in predicting improvement on the first post-test for the definition test. Participants with higher scores on the Michigan test trended towards being more likely to improve on the post-test, $\beta=0.19$, $SE(\beta)=0.10$, $z=1.94$, $p=.05$. Length of residence in an English-speaking country, gender, participant gender and participant age were not significant predictors of improvement on the first definition post-test.

1.2 Definition post-test 2

Mean score, percent (score out of 34), and percent change between pre-test and post-test 2 by group for the definition test is seen in Table 15. Mean percent change by group is presented in with 95% confidence interval standard error bars in Figure 22.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-Test</th>
<th>Post-Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score (SD)</td>
<td>Percent (SD)</td>
</tr>
<tr>
<td>Text-only</td>
<td>1.32 (2.24)</td>
<td>3.87 (6.58)</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>1.71 (2.37)</td>
<td>5.02 (6.96)</td>
</tr>
<tr>
<td>Text+Video</td>
<td>1.67 (2.74)</td>
<td>4.90 (8.07)</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>2.42 (2.32)</td>
<td>7.12 (6.81)</td>
</tr>
</tbody>
</table>

Table 15. Mean score, percent, and percent change by group for definition post-test 2 (n=34)
Table 15 above indicates the highest mean percent change for the Text+Audio+Video group (19.97%). However, the other multimedia groups, Text+Audio (14.71%) and Text+Video (15.03), did not outperform the Text-only group (16.10%). In order to determine whether these differences were statistically significant, a one-way between-subjects ANOVA was run and there was a non-significant main effect of group, $F(3, 69) = 0.69, p = .56, \eta^2 = .03$. In the definition post-test 2, none of the groups performed significantly better than the other groups.

A table of correlations between participant score and length of residence, vocabulary size, and Michigan test score by group can be seen in Table 16.
Table 16. Pearson’s Correlations between percent change and LOR, Michigan score, and vocabulary size on definition post-test 2

<table>
<thead>
<tr>
<th>Group</th>
<th>LOR</th>
<th>Michigan Score</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>.33</td>
<td>.50*</td>
<td>.46*</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>-.45†</td>
<td>.06</td>
<td>.25</td>
</tr>
<tr>
<td>Text+Video</td>
<td>.27</td>
<td>.44†</td>
<td>.33</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>-.16</td>
<td>.36</td>
<td>.07</td>
</tr>
</tbody>
</table>

† p<.10, * p<.05, ** p<.01, *** p<.001

A scatterplot of length of residence by percent difference between the definition pre-test and post-test 2 and length of residence by group with linear regression line can be seen in Figure 23.

A scatterplot of Michigan score by percent difference between the definition pre-test and post-test 2 and Michigan test score by group with linear regression line can be seen in Figure 24.
Figure 24. Scatterplot of participants’ percent difference between the definition pre-test and post-test two and Michigan score by group with linear regression line.

A scatterplot of vocabulary size by percent difference between the definition pre-test and post-test 2 and the vocabulary size by group with linear regression line can be seen in Figure 25.

Figure 25. Scatterplot of participants’ percent difference between the definition pre-test and post-test two and vocabulary size by group with linear regression line.

The analysis considered experimental variables of interest (i.e., instruction type and cue type) and other variables to control for any non-hypothesis related variance (i.e., item variable and subject variable). The logistic mixed effects model with random intercepts
for items and participants on the second definition post-test is presented below in Table 17. The model containing 13 predictor variables and two 2-way interactions was a significantly better fit than the base model without any predictor variables, $\chi^2(16)=52.44$, $p<.001$.

Table 17. Logistic mixed-effects model of participant improvement on definition post-test 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.77</td>
<td>0.47</td>
<td>-3.77</td>
<td>.000 ***</td>
</tr>
<tr>
<td><strong>Experimental Variables: Instruction Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruction.Audio</td>
<td>0.11</td>
<td>0.32</td>
<td>0.33</td>
<td>.744</td>
</tr>
<tr>
<td>Instruction.Video</td>
<td>0.12</td>
<td>0.33</td>
<td>0.37</td>
<td>.708</td>
</tr>
<tr>
<td>Instruction.Audio:Instruction.Video</td>
<td>0.13</td>
<td>0.46</td>
<td>0.28</td>
<td>.779</td>
</tr>
<tr>
<td><strong>Experimental Variables: Cue Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cue.Ling(ling)</td>
<td>1.11</td>
<td>0.40</td>
<td>2.75</td>
<td>.006 **</td>
</tr>
<tr>
<td>Cue.Vis(visual)</td>
<td>0.01</td>
<td>0.40</td>
<td>0.02</td>
<td>.986</td>
</tr>
<tr>
<td>Cue.Ling(ling):Cue.Vis(visual)</td>
<td>-0.17</td>
<td>0.55</td>
<td>-0.31</td>
<td>.759</td>
</tr>
<tr>
<td><strong>Item-level Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word.Frequency</td>
<td>-0.26</td>
<td>0.16</td>
<td>-1.66</td>
<td>.098 †</td>
</tr>
<tr>
<td>Word.Length</td>
<td>0.09</td>
<td>0.22</td>
<td>0.42</td>
<td>.677</td>
</tr>
<tr>
<td>PartOfSpeech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POS(verb)</td>
<td>-0.29</td>
<td>0.48</td>
<td>-0.62</td>
<td>.537</td>
</tr>
<tr>
<td>POS(adj)</td>
<td>-0.99</td>
<td>0.39</td>
<td>-2.56</td>
<td>.010 *</td>
</tr>
<tr>
<td>NumberOfWords(subtitle)</td>
<td>-0.46</td>
<td>0.24</td>
<td>-1.95</td>
<td>.051 †</td>
</tr>
<tr>
<td><strong>Subject-level Variables: Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender(M)</td>
<td>-0.44</td>
<td>0.25</td>
<td>-1.77</td>
<td>.077 †</td>
</tr>
<tr>
<td>Age</td>
<td>0.06</td>
<td>0.12</td>
<td>0.48</td>
<td>.631</td>
</tr>
<tr>
<td>LOR</td>
<td>-0.16</td>
<td>0.12</td>
<td>-1.36</td>
<td>.173</td>
</tr>
<tr>
<td>Vocab</td>
<td>0.41</td>
<td>0.14</td>
<td>2.87</td>
<td>.004 **</td>
</tr>
<tr>
<td>Michigan</td>
<td>0.34</td>
<td>0.14</td>
<td>2.41</td>
<td>.016 *</td>
</tr>
</tbody>
</table>

† $p<.10$, * $p<.05$, ** $p<.01$, *** $p<.001$

For the instructional variables of interest, participants who received audio instruction, the Text+Audio and Text+Audio+Video groups, were not significantly more likely to improve on the second definition post-test on average than participants who did not receive audio instruction, the Text-only and Text+Video groups, $\beta=0.11$, SE($\beta$)=0.32, $z=0.33$, $p=.74$. Participants who received video instruction, the Text+Video and Text+Audio+Video groups, were not significantly more likely to improve on the second definition post-test on average than participants who did not receive video instruction, the Text-only and Text+Audio groups, $\beta=0.12$, SE($\beta$)=0.33, $z=0.37$, $p=.71$. The two-way interaction of audio instruction and video instruction was not significant, $\beta=0.13$, SE($\beta$)=0.46, $z=-0.28$, $p=.78$.

For the cue variables of interest, participants were significantly more likely to improve on items which contained a linguistic cue than items without linguistic cues, $\beta=1.11$, SE($\beta$)=0.40, $z=2.75$, $p<.01$. Participants were not significantly more likely improve on items which contained visual cues than items without visual cues, $\beta=0.01$, SE($\beta$)=0.40, $z=0.02$, $p=.99$. The two-way interaction of linguistic cue and visual cue was not significant, $\beta=-0.17$, SE($\beta$)=0.55, $z=-0.31$, $p=.76$.

For the item-level variables, part of speech of the item significantly predicted improvement in the second definition post-test, while word frequency in English and number of words in the subtitle approached significance. Length of the word (i.e., the number of letters in the word) was not a significant predictor in modeling participant improvement on the post-test. Participants were significantly more likely to improve on nominal items than adjectival items ($\beta=-0.55$, SE($\beta$)=0.32, $z=-1.74$, $p=.08$) but were not significantly more likely to improve on nominal items than verbal items ($\beta=-0.25$,
SE(β)=0.39, z=-0.64, p=.53). Participants trended towards being less likely to improve on less frequent English words than more frequently used words, β=-0.26, SE(β)=0.16, z=-1.66, p=.10. Participants also trended towards being less likely to improve on items whose subtitle contain more words than items with less words, β=-0.46, SE(β)=0.24, z=-1.95, p=.05.

For the subject-level, demographic variables, vocabulary size and participant score on the Michigan test were significant predictors in modeling participant improvement. Participants’ gender approached significance in predicting improvement on the second post-test for the definition test. Length of residence in an English-speaking country and participant age were not significant predictors of improvement on the second definition post-test. The larger vocabulary size a participant had, the more likely they were to improve on post-test 2, β=0.41, SE(β)=0.14, z=2.87, p<.01. Participants with higher scores on the Michigan were significantly more likely to improve on post-test 2, β=0.34, SE(β)=0.14, z=2.41, p<.05. Male participants were significantly more likely to improve than female participants on second definition post-test, β=-0.44, SE(β)=0.25, z=-1.77, p=.08.

2. Multiple choice test analysis

2.1. Multiple choice post-test 1

Mean score, percent (score out of 34), and percent change between pre-test and post-test 1 by group for the multiple choice test is seen in Table 18. Mean percent change by group is presented with 95% confidence interval standard error bars in Figure 26.
Table 18. Mean score, percent, and percent change by group for multiple choice post-test 1 (n=34)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-Test</th>
<th>Post-Test 1</th>
<th>Post-Test 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score (SD)</td>
<td>Score (SD)</td>
<td>Percent Change (SD)</td>
</tr>
<tr>
<td>Text-only</td>
<td>4.95 (4.95)</td>
<td>19.68 (6.56)</td>
<td>43.34 (19.16)</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>9.59 (3.73)</td>
<td>25.24 (3.85)</td>
<td>46.02 (11.94)</td>
</tr>
<tr>
<td>Text+Video</td>
<td>5.22 (3.84)</td>
<td>21.44 (7.21)</td>
<td>47.71 (16.25)</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>6.63 (5.71)</td>
<td>24.21 (3.84)</td>
<td>51.70 (17.13)</td>
</tr>
</tbody>
</table>

Figure 26. Mean percent difference between the pre-test and post-test 1 for multiple choice questions with 95% confidence interval standard error bars by group

Table 18 above shows overall higher percent change for the multimedia groups, highest score for the Text+Audio+Video (51.70%) followed by Text+Video (47.71%) and Text+Audio (46.02%), compared to the Text-only group (43.34%). In order to determine
whether these differences were statistically significant, a one-way between-subjects ANOVA was run and there was a non-significant main effect of group, F(3,69)=0.86, p=.47, \( \eta^2=.04 \). In the multiple choice post-test 1, no one particular group significantly improved from the multiple choice pre-test than the other groups.

A table of correlations between participant score and LOR, vocabulary size, and Michigan score by group can be seen in Table 19.

Table 19. Pearson’s Correlations between percent change and LOR, MTELP score, and vocabulary size on multiple choice post-test 1

<table>
<thead>
<tr>
<th>Group</th>
<th>Length of Residence</th>
<th>Michigan Test Score</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>.04</td>
<td>.47*</td>
<td>.31</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>-.14</td>
<td>.24</td>
<td>.31</td>
</tr>
<tr>
<td>Text+Video</td>
<td>.08</td>
<td>.64**</td>
<td>.56*</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>-.52*</td>
<td>-.33</td>
<td>-.51*</td>
</tr>
</tbody>
</table>

* \( p<.05 \), ** \( p<.01 \), *** \( p<.001 \)

A scatterplot of length of residence by percent difference between the multiple choice pre-test and post-test 1 and length of residence by group with linear regression line can be seen in Figure 27.
Figure 27. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test one and LOR by group with linear regression line

A scatterplot of Michigan score by percent difference between the multiple choice pre-test and post-test 1 and Michigan score by group with linear regression line can be seen in Figure 28.

Figure 28. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 1 and Michigan score by group with linear regression line

A scatterplot of vocabulary size by percent difference between the multiple choice pre-test and post-test 1 and vocabulary size by group with linear regression line can be seen in Figure 29.
Figure 29. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test one and vocabulary size by group with linear regression line

The analysis considered experimental variables of interest (i.e., instruction type and cue type) and other variables to control for any non-hypothesis related variance (i.e., item variable and subject variable). The logistic mixed effects model with random intercepts for item and participant of participant improvement on the first multiple choice post-test is presented below in Table 20. The model containing 13 predictor variables and two 2-way interactions was a significantly better fit than the base model without any predictor variables, $\chi^2(16)=65.31, p<.001$. 
Table 20. Logistic mixed-effects model of participant improvement on multiple choice post-test 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.32</td>
<td>0.27</td>
<td>1.22</td>
<td>.222</td>
</tr>
</tbody>
</table>

**Experimental Variables: Instruction Type**

| Instruction.Audio | 0.13     | 0.18       | 0.72    | .471    |
| Instruction.Video | 0.19     | 0.18       | 1.03    | .302    |
| Instruction.Audio:Instruction.Video | 0.13     | 0.26       | 0.52    | .602    |

**Experimental Variables: Cue Type**

| Cue.Ling(ling) | 0.22     | 0.22       | 0.97    | .330    |
| Cue.Vis(visual) | 0.12     | 0.22       | 0.54    | .588    |
| Cue.Ling(ling):Cue.Vis(visual) | 0.06     | 0.30       | 0.20    | .845    |

**Item-level Variables**

| Word.Frequency | -0.07    | 0.09       | -0.84   | .402    |
| Word.Length    | 0.18     | 0.12       | 1.50    | .134    |
| PartOfSpeech   |          |            |         |         |
| POS(verb)      | -0.49    | 0.27       | -1.79   | .073    |
| POS(adj)       | -0.54    | 0.22       | -2.44   | .015    |
| NumberOfWords(subtitle) | -0.12    | 0.13       | -0.89   | .374    |

**Subject-level Variables: Demographics**

| Gender(M)       | -0.33    | 0.14       | -2.42   | .015    |
| Age             | 0.14     | 0.07       | 2.12    | .034    |
| LOR             | -0.07    | 0.07       | -1.07   | .286    |
| Vocab           | 0.20     | 0.08       | 2.72    | .006    |
| Michigan        | 0.26     | 0.08       | 3.28    | .001    |

† p<.10, * p<.05, ** p<.01, *** p<.001

**Formula in R**: Answer ~ Instruction.Audio + Instruction.Video + Instruction.Audio:Instruction.Video + Frequency + Cue.Ling + Cue.Vis + Cue.Ling:Cue.Vis + Length + POS + NumberOfWords + Gender + Age + LOR + Vocab + Michigan + (1 | Participant) + (1 | Item)

For the instructional variables of interest, participants who received audio instruction, the Text+Audio and Text+Audio+Video groups, were not significantly more likely to improve on the first multiple choice post-test on average than participants who did not receive audio instruction, the Text-only and Text+Video groups, β=0.13,
SE(β)=0.18, z=0.72, p=.47. Participants who received video instruction, the Text+Video and Text+Audio+Video groups, were not significantly more likely to improve on the first multiple choice post-test on average than participants who did not receive video instruction, the Text-only and Text+Audio groups, β=0.19, SE(β)=0.18, z=1.03, p=.30. The two-way interaction of audio instruction and video instruction was not significant, β=0.13, SE(β)=0.26, z=0.52, p=.60.

For the cue variables of interest, none of the variables were significant in modeling improvement on the first multiple choice post-test. Participants were not significantly more likely to improve on items which contained a linguistic cue than items without linguistic cues, β=0.22, SE(β)=0.22, z=0.97, p=.33. Participants were not significantly more likely to improve on items which contained visual cues than items without visual cues, β=0.12, SE(β)=0.22, z=0.54, p=.59. The two-way interaction of linguistic cue and visual cue was not significant, β=0.06, SE(β)=0.30, z=0.20, p=.86.

For the item-level variables, part of speech of the item was a significant predictor of participant improvement on the first multiple choice post-test. Word frequency in English, word length, and number of words in the subtitle were not significant predictors in modeling participant improvement on the post-test. Participants were significantly more likely to improve on nominal items than adjectival items (β=-0.54, SE(β)=0.22, z=-2.44, p<.05) and participants trended towards being more likely to improve on nominal items than verbal items (β=-0.49, SE(β)=0.27, z=-1.79, p=.07).

For the subject-level, demographic variables, participants’ gender, participants’ age, vocabulary size, and participant score on the Michigan test were significant predictor in modeling participant improvement on the first post-test for the multiple choice test.
Length of residence in an English-speaking country was not significant in modeling a participant’s improvement as measured by the first multiple choice post-test. Male participants were significantly less likely to improve than female participants on first multiple choice post-test, $\beta=-0.33$, SE($\beta)=0.14$, $z=-2.42$, $p<.05$. Older participants were significantly more likely to improve than younger participants, $\beta=0.14$, SE($\beta)=0.07$, $z=2.12$, $p<.05$. The larger vocabulary size a participant had, the more likely they were to improve on post-test, $\beta=0.20$, SE($\beta)=0.08$, $z=2.72$, $p<.01$. Participants with higher scores on the Michigan test of language proficiency were significantly more likely to improve on the post-test than participants with lower scores on the Michigan test of language proficiency, $\beta=0.26$, SE($\beta)=0.08$, $z=3.28$, $p<.01$.

2.2. **Multiple choice post-test 2**

Mean score, percent (score out of 34), and percent change between pre-test and post-test 2 by group for the multiple choice test is seen in Table 21. Mean percent change by group is presented with 95% confidence interval standard error bars in Figure 30.
Table 21. Mean score, percent, and percent change by group for multiple choice post-test 2 (n=34)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-Test</th>
<th>Post-Test 2</th>
<th>Post-Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score (SD)</td>
<td>Percent (SD)</td>
<td>Score (SD)</td>
</tr>
<tr>
<td>Text-only</td>
<td>4.95</td>
<td>14.55</td>
<td>16.47</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>9.59</td>
<td>28.20</td>
<td>17.29</td>
</tr>
<tr>
<td>Text+Video</td>
<td>5.22</td>
<td>15.36</td>
<td>16.06</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>6.63</td>
<td>19.50</td>
<td>17.42</td>
</tr>
</tbody>
</table>

Figure 30. Mean percent difference between the pre-test and post-test 2 for multiple choice questions with 95% confidence interval standard error bars by group

For multiple choice post-test 2, none of the multimedia groups improved as much as the Text-only group (33.90%), which scored the highest among the four groups. The Text+Audio group scored the lowest (22.66%) and Text+Video and Text+Audio+Video groups performed similarly (31.86% and 31.73% respectively). In order to determine
whether these differences were statistically significant, a one-way between-subjects ANOVA was run and there was a non-significant main effect of group, F(3,69)=1.60, p=.20, \(\eta^2=.07\). In the multiple choice post-test 2, no specific group revealed a significant improvement from the multiple choice pre-test.

A table of correlations between participant score and LOR, vocabulary size, and Michigan score by group can be seen in Table 22.

Table 22. Pearson’s Correlations between percent change and LOR, Michigan score, and vocabulary size on multiple choice post-test 2

<table>
<thead>
<tr>
<th>Group</th>
<th>LOR</th>
<th>Michigan Score</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>.06</td>
<td>.32</td>
<td>.19</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>.04</td>
<td>.43†</td>
<td>.25</td>
</tr>
<tr>
<td>Text+Video</td>
<td>-.06</td>
<td>.55*</td>
<td>.42†</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>-.55*</td>
<td>-.16</td>
<td>-.25</td>
</tr>
</tbody>
</table>

† p<.10, * p<.05, ** p<.01, *** p<.001

A scatterplot of length of residence by percent difference between the multiple choice pre-test and post-test 2 and length of residence by group with linear regression line can be seen in Figure 31.
Figure 31. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 2 and LOR by group with linear regression line

A scatterplot of Michigan score by percent difference between the multiple choice pre-test and post-test 2 and Michigan score by group with linear regression line can be seen in Figure 32.

Figure 32. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 2 and Michigan score by group with linear regression line

A scatterplot of vocabulary size by percent difference between the multiple choice pre-test and post-test 2 and vocabulary size by group with linear regression line can be seen in Figure 33.
Figure 33. Scatterplot of participants’ percent difference between the multiple choice pre-test and post-test 2 and vocabulary size by group with linear regression line

The analysis considered experimental variables of interest (i.e., instruction type and cue type) and other variables to control for any non-hypothesis related variance (i.e., item variable and subject variable). The logistic mixed effects model with random intercepts for item and participant of participant improvement on the second multiple choice post-test is presented below in Table 23. The model containing 13 predictor variables and two 2-way interactions was a significantly better fit than the base model without any predictor variables, \( \chi^2(16) = 43.73, p<.001 \).
Table 23. Logistic mixed-effects model of participant improvement on multiple choice post-test 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.35</td>
<td>0.32</td>
<td>-1.09</td>
<td>.275</td>
</tr>
</tbody>
</table>

**Experimental Variables: Instruction Type**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction.Audio</td>
<td>-0.18</td>
<td>0.21</td>
<td>-0.86</td>
<td>.390</td>
</tr>
<tr>
<td>Instruction.Video</td>
<td>-0.03</td>
<td>0.21</td>
<td>-0.16</td>
<td>.875</td>
</tr>
<tr>
<td>Instruction.Audio:Instruction.Video</td>
<td>0.32</td>
<td>0.30</td>
<td>1.06</td>
<td>.287</td>
</tr>
</tbody>
</table>

**Experimental Variables: Cue Type**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue.Ling(ling)</td>
<td>0.33</td>
<td>0.28</td>
<td>1.21</td>
<td>.226</td>
</tr>
<tr>
<td>Cue.Vis(visual)</td>
<td>0.18</td>
<td>0.27</td>
<td>0.66</td>
<td>.509</td>
</tr>
<tr>
<td>Cue.Ling(ling):Cue.Vis(visual)</td>
<td>0.13</td>
<td>0.37</td>
<td>0.34</td>
<td>.734</td>
</tr>
</tbody>
</table>

**Item-level Variables**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word.Frequency</td>
<td>-0.08</td>
<td>0.11</td>
<td>-0.71</td>
<td>.475</td>
</tr>
<tr>
<td>Word.Length</td>
<td>0.03</td>
<td>0.15</td>
<td>0.22</td>
<td>.823</td>
</tr>
<tr>
<td>PartOfSpeech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POS(verb)</td>
<td>-0.11</td>
<td>0.33</td>
<td>-0.33</td>
<td>.745</td>
</tr>
<tr>
<td>POS(adj)</td>
<td>-0.33</td>
<td>0.27</td>
<td>-1.20</td>
<td>.231</td>
</tr>
<tr>
<td>NumberOfWords(subtitle)</td>
<td>-0.14</td>
<td>0.16</td>
<td>-0.86</td>
<td>.389</td>
</tr>
</tbody>
</table>

**Subject-level Variables: Demographics**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender(M)</td>
<td>-0.49</td>
<td>0.16</td>
<td>-3.08</td>
<td>.002 **</td>
</tr>
<tr>
<td>Age</td>
<td>0.14</td>
<td>0.08</td>
<td>1.90</td>
<td>.058 †</td>
</tr>
<tr>
<td>LOR</td>
<td>-0.06</td>
<td>0.08</td>
<td>-0.71</td>
<td>.480</td>
</tr>
<tr>
<td>Vocab</td>
<td>0.20</td>
<td>0.09</td>
<td>2.28</td>
<td>.023 *</td>
</tr>
<tr>
<td>Michigan</td>
<td>0.19</td>
<td>0.09</td>
<td>2.02</td>
<td>.044 *</td>
</tr>
</tbody>
</table>

† p<.10, * p<.05, ** p<.01, *** p<.001

**Formula in R:** `Answer ~ Instruction.Audio + Instruction.Video + Instruction.Audio:Instruction.Video + Frequency + Cue.Ling + Cue.Vis + Cue.Ling:Cue.Vis + Length + POS + NumberOfWords + Gender + Age + LOR + Vocab + Michigan + (1 | Participant) + (1 | Item)`

For the instructional variables of interest, participants who received audio instruction, the Text+Audio and Text+Audio+Video groups, were not significantly less likely to improve on the second multiple choice post-test on average than participants who did not receive audio instruction, the Text-only and Text+Video groups, $\beta=-0.18$, 

99
SE(β)=0.21, z=-0.21, p=.39. Participants who received video instruction, the Text+Video and Text+Audio+Video groups, were not significantly less likely to improve on the second multiple choice post-test on average than participants who did not receive video instruction, the Text-only and Text+Audio groups, β=-0.03, SE(β)=0.21, z=-0.16, p=.86. The two-way interaction of audio instruction and video instruction was not significant, β=0.32, SE(β)=0.30, z=1.06, p=.29.

For the cue variables of interest, none of the variables were significant in modeling improvement on the second multiple choice post-test. Participants were not significantly more likely to improve on items which contained a linguistic cue than items without linguistic cues, β=0.33, SE(β)=0.28, z=1.21, p=.23. Participants were not significantly more likely to improve on items which contained visual cues than items without visual cues, β=0.18, SE(β)=0.27, z=0.66, p=.51. The two-way interaction of linguistic cue and visual cue was not significant, β=0.13, SE(β)=0.37, z=0.34, p=.73.

For the item-level variables, none of the variables were significant predictors in modeling participants’ improvement in learning English vocabulary items as measured by the second multiple choice post-test. Word frequency in English, word length, part of speech of the item, and number of words in the subtitle didn’t significantly contribute to the model of participants’ improvement.

For the subject-level, demographic variables, participants’ gender, participants’ age, vocabulary size, and participant score on the Michigan test were significant in predicting participant improvement on the second post-test for the multiple choice test. Length of residence in an English-speaking country was not significant in modeling a participants’ improvement as measured by the second multiple choice post-test. Male participants were
significantly less likely to improve than female participants on second multiple choice post-test, \( \beta = -0.49, SE(\beta) = 0.16, z = -3.08, p < .01 \). The larger vocabulary size a participant had, the more likely they were to improve on post-test 2, \( \beta = 0.20, SE(\beta) = 0.09, z = 2.28, p < .05 \). Participants with higher scores on the Michigan test were significantly more likely to improve on the second post-test, \( \beta = 0.19, SE(\beta) = 0.09, z = 2.02, p < .05 \). Older participants trended being significantly more likely to improve than younger participants, \( \beta = 0.14, SE(\beta) = 0.08, z = 1.90, p = .06 \).

3. **Post-experiment survey analysis**

   After each treatment, participants were asked two questions about the instruction they received in the questionnaire completed after the second post-test. The first question “How would you rate the instruction?” was included to evaluate participants’ attitudes towards instruction they received on a five-point scale (1=poor, 5=excellent). They were also asked a question “How useful was the instruction in remembering words that were taught?” to rate usefulness of the instruction in recalling the target words (1=don’t know, 5=very useful). Mean rating can be seen in Table 24 below. Instruction group difference for mean rating of instruction and mean rating of usefulness were explored using ANOVAs *post-hoc* t-test.
Table 24. Mean rating of instruction and usefulness by group

<table>
<thead>
<tr>
<th>Group</th>
<th>Rating of Instruction (SD)</th>
<th>Rating of Usefulness (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-only</td>
<td>3.74 (0.99)</td>
<td>3.47 (1.22)</td>
</tr>
<tr>
<td>Text+Audio</td>
<td>3.53 (0.72)</td>
<td>4.06 (0.66)</td>
</tr>
<tr>
<td>Text+Video</td>
<td>3.06aa (0.80)</td>
<td>4.21a (0.71)</td>
</tr>
<tr>
<td>Text+Audio+Video</td>
<td>4.21aa (0.92)</td>
<td>3.56a (0.62)</td>
</tr>
<tr>
<td>Total</td>
<td>3.64 (0.95)</td>
<td>3.82 (0.89)</td>
</tr>
<tr>
<td>F</td>
<td>5.62**</td>
<td>3.46*</td>
</tr>
</tbody>
</table>

Note: Within the predictor on both dependent variables, two categories share a common superscript if their difference is statistically significant at either .05, .01, or .001 level (“a” or “b” indicate *p<.05, “aa” or “bb” indicate *p<.01, and “aaa” or “bbb” represent **p<.001). Those compared means without a common superscript do not differ from each other at any of the levels of statistical significance considered.

Note: The F statistic is included and significance for the one-way ANOVA is indicated with an asterisk (* p<.05 ** p<.01 *** p<.001).

3.1. Instruction rating

A one-way ANOVA was run to see if there was a significant difference in post-experiment rating of instruction by group. The main effect of group was significant, F(3,69)=5.62, p<.01, η²=.20. Post-hoc pairwise t-tests with a Holm-Bonferroni were run to further explore the significant main effect of group. Mean instruction ratings for the Text-only group (M=3.74, SD=0.99) did not differ significantly from the mean instruction ratings for the Text+Audio group (M=3.53, SD=0.72; t(34)=0.71, p=.99, r=.12), neither from the Text+Video group (M=3.06, SD=0.80; t(35)=2.29, p=.17, r=.36), nor from the Text+Audio+Video group (M=4.21, SD=0.92; t(26)=1.53, p=.81, r=.25). Mean instruction ratings for the Text+Audio group did not differ significantly from the mean instruction ratings for neither the Text+Video group (t(33)=1.84, p=.45, r=.31), nor the text-audio-video group, t(34)=2.46, p=.11, r=.39. The only significant group difference was between the Text+Audio+Video group and the Text+Video group. The
Text+Audio+Video group rated the instruction in the course as significantly higher than the Text+Video group, \(t(35)=4.07, p<.01, r=.57\).

### 3.2. Usefulness rating

A one-way ANOVA was run to see if there was a significant difference in post-experiment rating of usefulness by group. The main effect of group was significant, \(F(3,69)=3.46, p<.05, \eta^2=.13\). Post-hoc pairwise t-tests with a Holm-Bonferroni were run to further explore the significant main effect of group. Mean instruction usefulness for the text only group (M=3.47, SD=1.22) did not differ significantly from the mean instruction ratings for the Text+Audio group (M=4.06, SD=0.66; \(t(28.28)=-1.82, p=.48, r=.32\)), neither from the Text+Video group (M=4.21, SD=0.71; \(t(26.94)=-0.26, p=.99, r=.05\)), nor from the Text+Audio+Video group (M=3.56, SD=0.62; \(t(29.04)=-2.27, p=.18, r=.39\)).

Mean instruction ratings for the Text+Audio group did not differ significantly from the mean instruction ratings for neither the Text+Video group \((t(33)=2.34, p=.15, r=.38)\), nor the Text+Audio+Video group, \(t(35)=-0.66, p=.99, r=.11\). The only significant group difference was between the Text+Audio+Video group and the Text+Video group. The Text+Audio+Video group rated the instruction in the course as significantly higher than the Text+Video group, \(t(35)=2.98, p<.05, r=.45\).
VIII. Summary of Results

The research questions of the present study were: (1) whether the use of multimedia presentation would facilitate L2 vocabulary acquisition of high-level, low-frequency English words as proposed by the dual-coding theory (i.e., Text-only vs. Text+Audio, Text+Video, Text+Audio+Video), (2) what type of multimedia presentation would best facilitate L2 vocabulary acquisition of high-level, low-frequency English words? (i.e., Text+Audio, Text+Video, Text+Audio+Video), (3) whether providing three modalities (i.e., Text+Audio+Video) would result in a lower retention level than two modalities (i.e., Text+Audio) as predicted by the cognitive load theory, and (4) what aspects of multimedia (i.e., linguistic cues vs. visual cues) would be the most effective in L2 vocabulary acquisition and retention of high-level, low-frequency English words. To this end, the study compared four different types of classroom instruction on high-level, low-frequency GRE English words among Korean learners of English: (1) Text-only (control group); (2) Text+Audio (no video); (3) Text+Video (no audio); and (4) Text+Audio+Video.

1. Research question 1

The first research question was whether the use of multimedia presentation would facilitate L2 vocabulary acquisition of high-level, low-frequency English words as proposed by the dual-coding theory (i.e., Text-only vs. Text+Audio, Text+Video, Text+Audio+Video). The dual-coding theory suggests that two symbolic systems, namely the imagery system and the verbal system, mediate cognitive activity (Clark & Paivio, 1991).
The percent changes from the definition pre-test to post-test 1 revealed positive tendencies towards multimedia instruction. There were significant effects of audio instruction and video instruction when using the logistic mixed effects model. Participants who received audio instruction, Text+Audio and Text+Audio+Video (M=45.00%), improved significantly more than those who did not receive audio instruction, Text-only and Text+Video (M=38.91%). Participants who received video instruction, Text+Video and Text+Audio+Video (M=44.33%) improved significantly more than those without video instruction, Text-only and Text+Audio (M=39.59%). The results from definition post-test 1 demonstrated that participants who had instruction with audio, video, or both benefited from instruction more than the Text-only group did, which appears to confirm multimedia effects supported by the body of literature (e.g., Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Hanley, Herron, & Cole, 1995; Kim & Gilman, 2008; Mohsen & Balakumar, 2011; Neuman & Koskinen, 1992; Schnotz, Böckheler, & Grzondziel, 1999; Yoshii & Flaitz, 2002).

Similar results did not replicate in the definition post-test 2, in which neither audio instruction nor video instruction showed significant effects on vocabulary acquisition. However, the Text+Audio+Video group (19.97%) still generated the highest percent change from the pre-test compared to other groups: Text-only (16.10%), Text+Video (15.03%), or Text+Audio (14.71%) although this group-level difference was not significant.

The improvement on percent changes for multiple choice post-test 1 was most salient in the Text+Audio+Video group (51.70%). All the multimedia groups,
Text+Audio, Text+Video, and Text+Audio+Video (M=48.48%), outperformed the Text-only group (43.34%); nonetheless, the effect was not significant.

The results from the multiple choice post-test 2 revealed a rather anomalous trend in that the Text+Audio group (22.66%) had the lowest percent change scores of all the other groups, Text-only (33.90%), Text+Video (31.86%), and Text+Audio+Video (31.73%). However, the effect of instruction was not significant.

In sum, there was a significant effect of multimedia instruction in the definition post-test 1 and positive trends towards Text+Audio+Video instruction across post-tests, which seems to endorse the dual-coding theory (Clark & Paivio, 1991) and the findings of the studies that substantiate benefits of multimedia components (e.g., Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Hanley, Herron, & Cole, 1995; Kim & Gilman, 2008; Mohsen & Balakumar, 2011; Neuman & Koskinen, 1992; Schnotz, Böckheler, & Grzondziel, 1999; Yoshii & Flaitz, 2002). The fact that percent changes from definition pre-test to definition post-test 1 were significant but this significance did not manifest in other post-tests may indicate that the multimedia effect needs to be sustained. Moreover, textual information was provided to all participants, including the Text-only group, and it is possible that participants whose English proficiency is rather advanced may have resorted to the text and learned the new words to a similar degree across groups. It is also possible that multiple choice questions are easier to guess correctly than definition questions because even when a student blindly guesses a correct answer, there is a 25% of the chance of being correct (Meara & Buxton, 1987), and this may have affected the results.
2. **Research question 2**

The second research question looked at what type of multimedia presentation would best facilitate L2 vocabulary acquisition of high-level, low-frequency English words (i.e., Text+Audio, Text+Video, Text+Audio+Video).

In the definition post-test 1, the Text+Audio+Video (45.20%) group outperformed the other two multimedia groups, Text+Audio (44.81%), and Text+Video (43.46%). The logistic mixed effects model showed significant effects of audio instruction and video instruction. The interaction of audio instruction and video instruction was not significant; therefore, the likelihood of improvement by the Text+Audio+Video group is best characterized by the simple addition of the effect of audio instruction and video instruction. To state another way, participants were more likely to improve with video instruction, as well as with audio instruction. Having a condition where the participants had both audio instruction and video instruction resulted in them outperforming either single-media group.

In the definition post-test 2, there seemed to be a strong tendency towards the Text+Audio+Video (19.97%) method as more effective than the rest of the multimedia groups: Text+Video, (15.03%) or Text+Audio (14.71%) although the effect of instruction was not significant.

Percent changes for the multiple choice post-test 1 showed the most positive outcome in the Text+Audio+Video group (51.70%), compared to Text+Video (47.71%) and Text+Audio (46.02%) although there was no significant effect of instruction.

Score improvement for the multiple choice post-test 2 suggested that the Text+Audio (22.66%) instruction was the least effective compared to the Text+Video
(31.86%) and the Text+Audio+Video group (31.73%). The logistic mixed effects model revealed a significantly negative effect of audio instruction. In these specific findings, it can be deduced that Text+Video or Text+Audio+Video instruction is more effective than Text+Audio instruction and that visual elements in multimedia presentation play an important role in expediting vocabulary learning.

To summarize, the results from the study exhibited significant effects of multimedia instruction in definition post-test 1, whether it is audio, video, or the combination of audio and video instruction. The Text+Audio+Video may be seen as the most effective multimedia method since the Text+Audio+Video group outperformed all the other multimedia groups across post-tests except the multiple choice post-test 2, where Text+Video group performed slightly better than the Text+Audio+Video group by 0.13%. The results from multiple choice post-test 2 also indicated that visual elements may have to be included in multimedia presentation for instruction to be more beneficial to learners as suggested by studies upholding video use for vocabulary learning (e.g., Baltova, 1999; Hanley, Herron, & Cole, 1995; Neuman & Koskinen, 1992).

3. **Research question 3**

The third research question was whether providing three modalities (i.e., Text+Audio+Video) would result in a lower retention level than two modalities (i.e., Text+Audio or Text+Video) as predicted by the *cognitive load theory*. To recapitulate, the *cognitive load theory* (Sweller, Van Merrienboer, & Paas, 1998) proposes that mixing two modes of presentation (e.g., audio and video) maximizes the working memory capacity and increases learning while three sources of information (e.g., text, audio, and
video) may cause a learner’s attention to be divided, decelerating the processing of given information.

There was a positive effect of audio and video instruction in the results from the definition post-test 1. However, when comparing two modalities, Text+Audio and Text+Video (M=44.14%) with three modalities, Text+Audio+Video (45.20%), there was no significant effect of the number of modalities. That is, the combination of the audio and video instruction seemed to be more effective than either audio or video alone but did not become more effective than the sum of the effects because the interaction of audio and video did not show a significant effect.

In the definition post-test 2, the fact that the Text+Audio+Video group (19.97%) outperformed the other multimedia groups, Text+Audio and Text+Video (M=14.87%), may suggest that three modalities do not weaken retention level as predicted by the cognitive load theory.

Percent changes for the multiple choice post-test 1 revealed a positive tendency towards three modalities, Text+Audio+Video instruction (51.70%), compared to two modalities, Text+Audio or Text+Video (M=46.87%). Yet, these results are merely trends as the logistic mixed-effects model displayed non-significant effect of instruction.

For the multiple choice post-test 2, the percent change was the least prominent in the Text+Audio group (22.66%), compared to Text+Video (31.86%) and Text+Audio+Video (31.73%). When comparing instruction with three modalities, Text+Audio+Video (31.73%), with that with two modalities, Text+Audio and Text+Video (M=27.26%), the mere number indicates that providing three modalities will not result in a lower retention level than two modalities.
To sum up, there was no significant effect of the number of modalities across post-tests. That is, instruction with three modalities, Text+Audio+Video, did not impede learning process when compared with instruction with two modalities, Text+Audio or Text+Video. Therefore, findings did not confirm the cognitive load theory (Sweller, Van Merrienboer, & Paas, 1998) and suggested instead that mixing three modalities would not decelerate the processing of information.

4. Research question 4

The last research question was what aspect in multimedia (i.e., linguistic cues vs. visual cues) would be the most effective in L2 vocabulary acquisition and retention of high-level, low-frequency English words.

In the definition post-test 1, participants improved significantly more on words with a linguistic cue than words without a linguistic cue. Linguistic cues were significant in predicting participant improvement whereas visual cues did not play a significant role in facilitating learning of the test items. The interaction between the two cue types (i.e., linguistic+visual) was not significant, either.

In the definition post-test 2, similar results were found when looking at participant improvement on words containing a linguistic cue versus words without a linguistic cue. The linguistic cue type was significant in predicting participant improvement. Participants improved significantly more on words with a linguistic cue than words without a linguistic cue. The interaction between the linguistic cue and the visual cue was not significant.
In the multiple choice post-test 1, there was no significant effect on cue types. Participants did not significantly improve more on words with a linguistic cue than on words without a linguistic cue. Likewise, participants did not significantly more on words with a visual cue than on words without a visual cue.

Similar non-significant results were found in the multiple choice post-test 2: participants did not improve significantly more on words with a linguistic cue than words without a linguistic cue. Nor did they more significantly improve on words with a visual cue than words without a visual cue.

To summarize, results from definition post-test 1 and definition post-test 2 indicated significant effect of cue variables (i.e. linguistic vs. visual). The participants did improve significantly more on the words whose subtitles included a linguistic cue (e.g., synonyms or definitions) than on those without a linguistic cue (i.e., visual cue or no cue). The results from the definition post-tests were in alignment with studies that support the use of an annotation, such as a brief definition or explanation of a word (e.g., Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Hulstijn, 1992; Hulstijn, Hollander, & Greidanus 1996; Jones, 2004).

5. Residual results

In addition to the experimental variables (i.e., instruction type and cue-type), item-level variables and subject-level variables were also analyzed. The item-variables consisted of word frequency, word length, part of speech (POS), and number of words in the subtitle. The subject-level variables looked at gender, age, length of residence (LOR) in an English-speaking country, the Vocabulary Size Test, and the Michigan test.
For the item-variables, POS of the item showed the most significant effect across most post-tests. It approached significance in the definition post-test 1 and significantly predicted improvement in the definition post-test 2 and multiple choice post-test 1. In all of these tests, participants performed better on nominal items than adjectival items. Word frequency and the length of subtitles approached significance only in the definition post-test 2. Less frequent words with longer subtitles showed less improvement than were more frequent words with shorter subtitles. However, in both the multiple choice post-tests, word frequency in English, word length, and number of words in the subtitle did not significantly contribute to the model of participants’ improvement.

At the subject-level, vocabulary size had the most significant effect across post-tests. For all the post-tests, the larger the vocabulary size participants had, the more they improved on each post-test. The Michigan test mostly showed significance in predicting improvement across post-tests. Participants with a higher score on the Michigan test improved more on each post-test. Participants’ gender and age were significant in predicting improvement on multiple choice post-tests. Male participants improved significantly less than female participants and older participants improved significantly more than younger participants on the multiple choice post-tests. However, length of residence in an English-speaking country was not significant in modeling a participant’s improvement on low-frequency words in any of the post-tests.
IX. Discussion and Conclusions

The results from the definition post-test 1 indicated that multimedia presentation was indeed more advantageous than mere textual presentation, supporting the dual-coding theory. As suggested by the dual-coding theory (Clark & Paivio, 1991; Mayer, 1997), learning was promoted when utilizing multiple sensory modalities with visual and verbal cues. The results are also in line with the body of literature that advocates multimedia effects (e.g., Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Hanley Herron, & Cole, 1995; Kim & Gilman, 2008; Mohsen & Balakumar, 2011; Neuman & Koskinen, 1992; Schnottz, Böckheler, & Grzondziel, 1999; Yoshii & Flaitz, 2002). In fact, except for the results from the multiple choice post-test 2, in which the Text-only group outperformed all the other groups, percent changes were higher for multimedia groups, most prominently for the Text+Audio+Video group.

In all the post-tests except for the multiple choice post-test 2, there was a tendency towards the Text+Audio+Video method as more effective than Text+Audio or Text+Video. Further, results of the multiple choice post-test 2 may suggest Text+Audio is the least effective multimedia method. In comparing two modalities, Text+Audio or Text+Video, with three modalities, Text+Audio+Video, there was no significant effect of the number of modalities. On the contrary, the combination of the audio and video instruction seems to be more effective than either audio or video alone. Accordingly, the study did not support the cognitive load theory (Sweller, Van Merrienboer, & Paas, 1998), which suggests that three sources of information may overload learners’ cognitive resources and decelerate the processing of information, and favors two modes of presentation over three sources of information.
With regard to multimedia presentation, and what it should include, results from both the definition post-test 1 and 2 indicated significant effects of linguistic cues, which seems to prove the value of linguistic support in the elaboration of meaning of an unknown word. This has also been asserted by various other researchers (e.g., Hulstijn, 1992; Hulstijn, Hollander, & Greidanus, 1996; Jones, 2004). Although Nation (1982) and Nation and Coady (1988) indicated that learners learn new words by inferring their meaning from the context, it is also possible that learners infer an incorrect meaning of an unfamiliar word when not knowing its exact meaning (Hulstijn, 1992). The benefits of word definitions were exhibited in several studies (e.g., Jones, 2004; Kim & Gilman, 2008; Plass, Chun, Mayer, & Leutner, 1998), which suggested that when an unknown word is shown by itself without its meaning or in a context where the meaning is not interpreted, the given textual information may not be particularly advantageous to learners (Hulstijn, 1992). In contrast, providing learners with the definition of new words may accelerate learners’ understanding of the unknown word and enable learners to remember its meaning within a relatively short period of time. Therefore, the results of the present study are grounds for the recommendation that multimedia visual presentation include linguistic cues, such as the precise meaning of an unknown word to help learners acquire unknown L2 words.

The fact that delayed post-test results did not show instruction effect may suggest that multimedia support must be sustained to have an effect, which could be due to the nature of low-frequency test items. As some researchers affirm (e.g., Laufer, 2006; Laufer & Nation, 2011), an unknown word should be encountered at least ten times in order for a learner to have a chance of recognizing the meaning of the word. Considering
the difficulty and infrequent use of the test items, the current experiment exposed the participants to the test items only twice and therefore might not have provided the learners with enough exposure to the new words.

The present study investigated effects of multimedia instruction on high-level, low-frequency GRE words among Korean learners of English. For the most part, the study revealed positive effects of multimedia use in L2 vocabulary instruction. The findings of the study support the use of multimedia aids, most preferably in the form of Text+Audio+Video. Offering a visual component in addition to audio appears to be essential in multimedia instruction as the Text+Audio group displayed the least progress among the multimedia conditions. For multimedia to have sustained benefits, it is recommended that teachers offer multimedia presentation on numerous occasions for better outcomes, especially when teaching high-level, low-frequency words. As far as aspects of multimedia presentation and support are concerned, linguistic cues are found to be powerful in the retention of new L2 words; that is, a precise definition of an unknown word in a video enhances multimedia support and expedites acquisition of new L2 vocabulary. The findings of this study have both theoretical and pedagogical implications in L2 vocabulary acquisition of infrequent advanced-level vocabulary words. The study also contributes to L2 vocabulary teaching by addressing ways to design multimedia materials that can promote meaningful learning and by offering instructional guidelines for multimedia language teaching.
X. Appendices

Appendix A. Test items with subtitles

(LVC: linguistic and visual cues, LC: linguistic cue, VC: visual cue, NC: no cue)

<table>
<thead>
<tr>
<th>Word</th>
<th>Subtitle</th>
<th>Video type</th>
<th>Time</th>
<th>Freq.</th>
<th># of words</th>
<th>POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. beguiling</td>
<td>There’s also the engagement of my son Daniel to the lovely and beguiling Miss Emily Thorn.</td>
<td>LC</td>
<td>00:07</td>
<td>33,008</td>
<td>16</td>
<td>adj.</td>
</tr>
<tr>
<td>2. coax</td>
<td>A: Marie, honey. I need you to come out of there. B: No. A: Could you go in there and coax her out.</td>
<td>NC</td>
<td>00:06</td>
<td>27,078</td>
<td>20</td>
<td>verb</td>
</tr>
<tr>
<td>3. commotion</td>
<td>A: You know that he would. You know that. B: What’s all the commotion?</td>
<td>NC</td>
<td>00:04</td>
<td>19,254</td>
<td>11</td>
<td>noun</td>
</tr>
<tr>
<td>4. convivial</td>
<td>A: You’re gonna ruin it. B: Lily, I promise I won’t ruin your party. I’ll be charming and convivial.</td>
<td>LC</td>
<td>00:06</td>
<td>32,231</td>
<td>17</td>
<td>adj.</td>
</tr>
<tr>
<td>5. exasperation</td>
<td>Ugh. The audible sigh is a show of exasperation, right? Right.</td>
<td>LVC</td>
<td>00:09</td>
<td>15,255</td>
<td>11</td>
<td>noun</td>
</tr>
<tr>
<td>6. flinch</td>
<td>Robin fixes her hair. I flinch, I am doing so much flinching. It’s bad for my skin.</td>
<td>VC</td>
<td>00:04</td>
<td>23,764</td>
<td>17</td>
<td>verb</td>
</tr>
<tr>
<td>7. fluster</td>
<td>I’m a lawyer. I argue for a living, but when it comes to Lily, I just get all flustered.</td>
<td>NC</td>
<td>00:06</td>
<td>84,353</td>
<td>19</td>
<td>verb</td>
</tr>
<tr>
<td>8. garbled</td>
<td>A: Daddy, I can’t read this. Could you please read this for me? B: Which might explain the garbled note you told me about?</td>
<td>LVC</td>
<td>00:10</td>
<td>36,799</td>
<td>22</td>
<td>adj.</td>
</tr>
<tr>
<td>9. gloat</td>
<td>A: I thought I’d stop by and say hi. B: You mean stop by and gloat. A: Maybe a little.</td>
<td>NC</td>
<td>00:06</td>
<td>46,926</td>
<td>17</td>
<td>verb</td>
</tr>
<tr>
<td>10. gobble</td>
<td>She gobbled up hot wings and swallowed the bones.</td>
<td>LVC</td>
<td>00:04</td>
<td>39,890</td>
<td>9</td>
<td>verb</td>
</tr>
<tr>
<td>11. gossamer</td>
<td>Pulled out the gossamer curtains and found out…</td>
<td>VC</td>
<td>00:03</td>
<td>46,065</td>
<td>8</td>
<td>adj.</td>
</tr>
<tr>
<td>12. grovel</td>
<td>A: King, you’re majesty. I gravel at your feet. B: It’s not gravel, it’s grovel.</td>
<td>VC</td>
<td>00:08</td>
<td>50,618</td>
<td>13</td>
<td>verb</td>
</tr>
<tr>
<td>13. hubris</td>
<td>I did this. Monty was killed by my hubris and my pride.</td>
<td>LC</td>
<td>00:06</td>
<td>39,074</td>
<td>12</td>
<td>noun</td>
</tr>
</tbody>
</table>
| 14. lugubrious | A: Is everything okay with you?  
B: Me? Why?  
A: You seem a little **lugubrious**. | NC | 00:11 | 34,884 | 12 | adj. |
| 15. miff | A: The truth is, John, as senior partner, it doesn’t make a lot of sense for you to be dating associates.  
B: You seem a little **miffed** there. | NC | 00:10 | 38,885 | 26 | verb |
| 16. nonchalant | A: That’s Ross. What are we gonna do?  
B: Just be calm. Just be calm. For all he knows, we’re just hanging out. So just be **nonchalant**. | LVC | 00:16 | 36,471 | 25 | adj. |
| 17. pamper | You’re privileged. You’re **pampered**. You’re spoiled. You want what you want when you want it. You want excitement. | LC | 00:08 | 51,865 | 18 | verb |
| 18. perturbed | A: Sounds like you’re a little **perturbed** about that.  
B: I’m a little perturbed. I mean, it’s been three weeks. I haven’t had a good meal, and I don’t have clean underwear. | LVC | 00:11 | 23,848 | 30 | verb |
| 19. petulant | He also thinks you’re **petulant**, whiny, spoiled brat just like I was. | LC | 00:04 | 30,311 | 12 | adj. |
| 20. pristine | Clean water, fresh air, and **pristine** wilderness. | LVC | 00:02 | 19,109 | 7 | adj. |
| 21. pulchritude | A: As you can see, I don’t look like that. That was a moment of youthful **pulchritude** that is long since passed.  
B: Youthful **pulchritude**?  
A: Don’t ask me what **pulchritude** means.  
B: **Pulchritude** means beauty. | LC | 00:15 | NA | 32 | noun |
| 22. repugnant | Every woman, no matter how initially **repugnant**… | VC | 00:06 | 29,720 | 7 | adj. |
| 23. smug | A: Still looking well.  
B: Don’t be **smug**, Jacky. | NC | 00:04 | 16,283 | 7 | adj. |
| 24. solder | A: What the hell’s going on?  
B: I’m **soldering**. | VC | 00:03 | 28,131 | 7 | verb |
| 25. squabble | Alright, that’s enough juvenile **squabbling**, you stop it. You stop it, I say. | VC | 00:04 | 33,542 | 13 | verb |
| 26. squalid | ‘Cause I live at Central Park west, and you probably live at some **squalid** little studio apartment. | NC | 00:05 | 22,842 | 17 | adj. |
| 27. staunchly | And thank you for defending me so **staunchly**, Nell. | NC | 00:03 | 35,488 | 9 | adv. |
| 28. stymie | A: Be a lamb and open it for me.  
B: Why? What's the problem?  
A: You think you have me *stymied*, don't you? | VC | 00:13 | 61,218 | 20 | verb |
| 29. swerve | But, Angie, I’m telling you. This guy, he drives past me, he turns around, and *swerves* right into me. | VC | 00:10 | 36,390 | 19 | verb |
| 30. torment | Young tyrants who bring pain, intimidation and violence. What can children do in face of such *torment*? | LVC | 00:12 | 14,641 | 17 | noun |
| 31. toupee | Maybe a wig of some sort, or *toupee*? Does anyone wear a *toupee*? | LVC | 00:11 | 60,872 | 13 | noun |
| 32. vitiate | I will *vitiate* that prenuptial contract as sure as I’m saying here. I will break that agreement. If I don’t do it with a judge, I will do that with a jury and you consider that prospect, sir. | LC | 00:09 | 60,014 | 38 | verb |
| 33. wanton | It involved a long-standing client, which I also lost. I’ve lost other business all because he is *wanton*, predatory. Unwelcomed sexual conduct. | LC | 00:12 | 23,826 | 22 | adj. |
| 34. yank | A: Now, you hold still, and I’ll *yank* this thing out.  
B: Easy with the *yanking*. | VC | 00:05 | 36,030 | 16 | verb |
Appendix B. *Vocabulary definition*

<table>
<thead>
<tr>
<th>Word</th>
<th>IPA symbol</th>
<th>Definition (Korean)</th>
<th>Definition (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>beguiling</td>
<td>/bigáil̂/</td>
<td>adj. 매혹적인, 현혹하는</td>
<td>charming or enchanting, sometimes in a deceptive way</td>
</tr>
<tr>
<td>coax</td>
<td>/kouks/</td>
<td>v. (남을) 말래서 ~하도록 하다, 부추기다</td>
<td>to persuade or try to persuade by pleading or flattery</td>
</tr>
<tr>
<td>commotion</td>
<td>/kəmˈɒfjən/</td>
<td>n. 동요, 종분: 소란</td>
<td>a state of confused and noisy disturbance</td>
</tr>
<tr>
<td>convivial</td>
<td>/kənˈvɪvɪəl/</td>
<td>adj. 친목적인; 명랑한, 우호적인</td>
<td>cheerful and friendly; sociable</td>
</tr>
<tr>
<td>exasperation</td>
<td>/ɪɡˈzɛspərəʃən/</td>
<td>n. 분개, 저노.</td>
<td>the feeling of intense irritation or annoyance</td>
</tr>
<tr>
<td>flinch</td>
<td>/flɪntʃ/</td>
<td>v. 주춤하다, 움찔하다, 꽁무니 빼다</td>
<td>to make a quick, nervous movement of the face or body as an instinctive reaction to surprise, fear or pain</td>
</tr>
<tr>
<td>fluster</td>
<td>/flʌstər/</td>
<td>v. 혼란시키다, 당황케 하다</td>
<td>to make agitated or confused</td>
</tr>
<tr>
<td>garble</td>
<td>/ɡə.rəbl/</td>
<td>v. (사실, 서류 등) 왜곡하다. 오해하다.</td>
<td>to reproduce (a message, sound, or transmission) in a confused and distorted way</td>
</tr>
<tr>
<td>gloat</td>
<td>/ɡlɔut/</td>
<td>v. 만족스럽게, 자랑스럽게 보다; (자신의 성공에) 훌륭해 하다; (남의 실패를) 고소해 하다</td>
<td>to feel or express great, often malicious, pleasure or self-satisfaction</td>
</tr>
<tr>
<td>gobble</td>
<td>/ɡəbəl/</td>
<td>v. 급하게, 거칠게 먹다</td>
<td>to eat hurriedly and noisily</td>
</tr>
<tr>
<td>gossamer</td>
<td>/ɡəsəmər/</td>
<td>adj. 얇고 가벼운</td>
<td>sheer, light, delicate, or tenuous</td>
</tr>
<tr>
<td>grovel</td>
<td>/ɡrʌvəl/</td>
<td>v. 비꿀하게 굴다; 얻들러 기다</td>
<td>to behave in a servile or demeaning manner; to lie or creep in a prostrate position</td>
</tr>
<tr>
<td>hubris</td>
<td>/hjǔ.bris/</td>
<td>n. 자신과의: 과만</td>
<td>excessive pride or self-confidence</td>
</tr>
<tr>
<td>lugubrious</td>
<td>/luːˈɡʌbrɪəs/</td>
<td>adj. (과정더게) 슬픈, 유혈한</td>
<td>looking or sounding sad and dismal</td>
</tr>
<tr>
<td>miffed</td>
<td>/mɪft/</td>
<td>adj. 발끈한</td>
<td>offended or annoyed</td>
</tr>
<tr>
<td>nonchalant</td>
<td>/nʌnˈʃoʊlənt/</td>
<td>adj. 태연한, 차분한; 무관심한, 냉담한</td>
<td>feeling or appearing casually calm and relaxed</td>
</tr>
<tr>
<td>Word</td>
<td>Pronunciation</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>pamper</td>
<td>/'pæmper/</td>
<td>v. 에게 즐지하다; (육체적) 만족시키다 to indulge with every attention, comfort, and kindness; to spoil</td>
<td></td>
</tr>
<tr>
<td>perturbed</td>
<td>/ˈpɔrtəd/</td>
<td>adj. 혼란스러운; 당황한, 불안한 anxious or unsettled; upset</td>
<td></td>
</tr>
<tr>
<td>petulant</td>
<td>/ˈpetələnt/</td>
<td>adj. 불안정한, 초조해 하는 unreasonably irritable or ill-tempered</td>
<td></td>
</tr>
<tr>
<td>pristine</td>
<td>/ˈprɪstɪn/</td>
<td>adj. 초기 상태의; 오염되지 않은 in its original condition; unspoiled</td>
<td></td>
</tr>
<tr>
<td>pulchritude</td>
<td>/ˈpʌlkrɪtjuːd/</td>
<td>n. 미모, 육체미 great physical beauty and appeal</td>
<td></td>
</tr>
<tr>
<td>repugnant</td>
<td>/ˈrɪpəgnənt/</td>
<td>adj. 싫은, 불쾌한 extremely distasteful</td>
<td></td>
</tr>
<tr>
<td>smug</td>
<td>/smʌɡ/</td>
<td>adj. 자부심이 많은, 젖난 체하는; 독신적인 having or showing an excessive pride in oneself or one’s achievements</td>
<td></td>
</tr>
<tr>
<td>solder</td>
<td>/ˈsædər/</td>
<td>v. 납땜질하다; 결합하다 to join with a low-melting alloy</td>
<td></td>
</tr>
<tr>
<td>squabble</td>
<td>/ˈskwæbəl/</td>
<td>v. 쓸데없는 싸움, 말다툼을 하다 to quarrel noisily over a trivial matter</td>
<td></td>
</tr>
<tr>
<td>squalid</td>
<td>/ˈskwɔːld/</td>
<td>adj. 지저분한, 더러운, 누추한 extremely dirty and unpleasant, especially as a result of poverty or neglect</td>
<td></td>
</tr>
<tr>
<td>staunch</td>
<td>/stɔːnt/</td>
<td>adj. 견실한, 충실한, 확고한 loyal and committed in attitude</td>
<td></td>
</tr>
<tr>
<td>stymie</td>
<td>/ˈstʌmi/</td>
<td>v. 방해하다, 좌절시키다 to prevent or hinder the progress of</td>
<td></td>
</tr>
<tr>
<td>swerve</td>
<td>/ˈswɑːrv/</td>
<td>v. 갑자기 뛰어나가다, 빗어나다, 이탈하다 to change or cause to change direction abruptly</td>
<td></td>
</tr>
<tr>
<td>torment</td>
<td>/ˈtɔːmt/</td>
<td>n. (육체적, 정신적인) 고통, 고통 severe physical or mental suffering</td>
<td></td>
</tr>
<tr>
<td>toupee</td>
<td>/tuːˈpɛi/</td>
<td>n. (머리를 덮는) 부분 가발 a partial wig or hairpiece worn to cover a bald spot</td>
<td></td>
</tr>
<tr>
<td>vitiate</td>
<td>/ˈvɪʃiət/</td>
<td>v. ~의 결을 손상시키다, ~을 방치하다, 오염시키다 to spoil or impair the quality or efficiency of</td>
<td></td>
</tr>
<tr>
<td>wanton</td>
<td>/ˈwɒtən/</td>
<td>adj. 부당한, 터무니없는, 재밌다고; 바람기 있는, 호색적인 deliberate or unprompted; sexually immodest or promiscuous</td>
<td></td>
</tr>
<tr>
<td>yank</td>
<td>/jæŋk/</td>
<td>v. 힘을 가해 당기다 to pull with a quick, strong movement</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C. Examples of pre- and post-tests: Definition (n=40; distractors*)

1. solder ______________________
2. convivial ____________________
3. beguiling ____________________
4. coax _________________________
5. swerve ________________________
6. exasperation _________________
7. flinch _________________________
8. staunchly ______________________
9. garbled ________________________
10. wanton _______________________
11. gossamer _____________________
12. grovel _________________________
13. hubris _________________________
14. squabble _______________________
15. lugubrious _____________________
16. miffed _________________________
17. pristine ________________________
18. besmirch* ______________________
19. pamper _________________________
20. vindicate* _____________________
21. pulchritude ____________________
22. repugnant _____________________
23. smug _________________________
24. fluster _________________________
25. squalid _________________________
26. commotion _____________________
27. stymie _________________________
28. gobble _________________________
29. torment _________________________
30. toupee _________________________
31. gloat _________________________
32. vitiate _________________________
33. yank _________________________
34. banter* _________________________
35. nonchalant _____________________
36. perturbed _______________________
37. gripe* _________________________
38. fling* _________________________
39. strut* _________________________
40. petulant _________________________
Appendix D. *Examples of pre- and post-tests: Multiple choice (n=40; distractors*)*

1. **beguiling**
   - (a) disgusting
   - (b) charming
   - (c) confusing
   - (d) exciting

2. **commotion**
   - (a) disturbance
   - (b) cooperation
   - (c) accommodation
   - (d) expansion

3. **repugnant**
   - (a) distasteful
   - (b) superior
   - (c) critical
   - (d) greedy

4. **gossamer**
   - (a) casual
   - (b) delicate
   - (c) chatty
   - (d) mysterious

5. **exasperation**
   - (a) annoyance
   - (b) luxuriance
   - (c) isolation
   - (d) termination

6. **squalid**
   - (a) dirty
   - (b) strong
   - (c) little
   - (d) dizzy

7. **fluster**
   - (a) flatter
   - (b) agitate
   - (c) gather
   - (d) prosper

8. **garble**
   - (a) rinse
   - (b) deposit
   - (c) swallow
   - (d) distort

9. **fling**
   - (a) spark
   - (b) move
   - (c) throw
   - (d) fail

10. **coax**
    - (a) praise
    - (b) deceive
    - (c) persuade
    - (d) criticize

11. **gloat**
    - (a) express pleasure
    - (b) eat noisily
    - (c) torture
    - (d) corrupt

12. **gobble**
    - (a) work superficially
    - (b) talk sweetly
    - (c) eat hurriedly
    - (d) fight verbally
13. **staunch**
   (a) fat  
   (b) loyal  
   (c) original  
   (d) smelly  

14. **hubris**
   (a) pride  
   (b) importance  
   (c) inflexibility  
   (d) piece  

15. **lugubrious**
   (a) foolish  
   (b) vivid  
   (c) careless  
   (d) sorrowful  

16. **miffed**
   (a) limited  
   (b) annoyed  
   (c) complicated  
   (d) excited  

17. **nonchalant**
   (a) fancy  
   (b) inexpensive  
   (c) calm  
   (d) difficult  

18. **pamper**
   (a) spoil  
   (b) prohibit  
   (c) touch  
   (d) blame  

19. **perturbed**
   (a) damaged  
   (b) accepted  
   (c) clean  
   (d) anxious  

20. **solder**
   (a) pollute  
   (b) harden  
   (c) conjoin  
   (d) recruit  

21. **petulant**
   (a) arrogant  
   (b) ill-tempered  
   (c) flowery  
   (d) disgusting  

22. **yank**
   (a) pull fast  
   (b) beat up  
   (c) harass  
   (d) sleep tight  

23. **pristine**
   (a) excited  
   (b) religious  
   (c) unspoiled  
   (d) expensive  

24. **flinch**
   (a) force greatly  
   (b) hurt seriously  
   (c) move quickly  
   (d) interrupt suddenly  

25. **pulchritude**
   (a) pollution  
   (b) honesty  
   (c) arrogance  
   (d) beauty  

26. **convivial**
   (a) violent  
   (b) mutual  
   (c) minor  
   (d) lively
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>27. <strong>smug</strong></td>
<td>34. <strong>toupee</strong></td>
</tr>
<tr>
<td>(a) aggressive</td>
<td>(a) group</td>
</tr>
<tr>
<td>(b) proud</td>
<td>(b) trap</td>
</tr>
<tr>
<td>(c) hesitant</td>
<td>(c) alarm</td>
</tr>
<tr>
<td>(d) upsetting</td>
<td>(d) wig</td>
</tr>
<tr>
<td>28. <strong>squabble</strong></td>
<td>35. <strong>vitiate</strong></td>
</tr>
<tr>
<td>(a) move</td>
<td>(a) energize</td>
</tr>
<tr>
<td>(b) agree</td>
<td>(b) impair</td>
</tr>
<tr>
<td>(c) scratch</td>
<td>(c) whisper</td>
</tr>
<tr>
<td>(d) quarrel</td>
<td>(d) facilitate</td>
</tr>
<tr>
<td>29. <strong>stymie</strong></td>
<td>36. <strong>grovel</strong></td>
</tr>
<tr>
<td>(a) imprison</td>
<td>(a) dig</td>
</tr>
<tr>
<td>(b) irritate</td>
<td>(b) bark</td>
</tr>
<tr>
<td>(c) prevent</td>
<td>(c) develop</td>
</tr>
<tr>
<td>(d) satisfy</td>
<td>(d) crawl</td>
</tr>
<tr>
<td>30. <strong>wanton</strong></td>
<td>37. <strong>banter</strong>*</td>
</tr>
<tr>
<td>(a) fragile</td>
<td>(a) joke</td>
</tr>
<tr>
<td>(b) colorful</td>
<td>(b) stick</td>
</tr>
<tr>
<td>(c) promiscuous</td>
<td>(c) curve</td>
</tr>
<tr>
<td>(d) desirable</td>
<td>(d) prohibition</td>
</tr>
<tr>
<td>31. <strong>swerve</strong></td>
<td>38. <strong>gripe</strong>*</td>
</tr>
<tr>
<td>(a) avoid deliberately</td>
<td>(a) hold</td>
</tr>
<tr>
<td>(b) display emotion</td>
<td>(b) crush</td>
</tr>
<tr>
<td>(c) change direction</td>
<td>(c) smile</td>
</tr>
<tr>
<td>(d) present visually</td>
<td>(d) complain</td>
</tr>
<tr>
<td>32. <strong>torment</strong></td>
<td>39. <strong>besmirch</strong>*</td>
</tr>
<tr>
<td>(a) intelligence</td>
<td>(a) dishonor</td>
</tr>
<tr>
<td>(b) suffering</td>
<td>(b) surround</td>
</tr>
<tr>
<td>(c) endurance</td>
<td>(c) bow</td>
</tr>
<tr>
<td>(d) failing</td>
<td>(d) present</td>
</tr>
<tr>
<td>33. <strong>strut</strong>*</td>
<td>40. <strong>vindicate</strong>*</td>
</tr>
<tr>
<td>(a) beat</td>
<td>(a) revenge</td>
</tr>
<tr>
<td>(b) parade</td>
<td>(b) disturb</td>
</tr>
<tr>
<td>(c) fasten</td>
<td>(c) confirm</td>
</tr>
<tr>
<td>(d) arrange</td>
<td>(d) justify</td>
</tr>
</tbody>
</table>
Appendix E. Questionnaire

1. Which group do you belong to?
   - TEXT-ONLY
   - TEXT+AUDIO
   - TEXT+VIDEO
   - TEXT+AUDIO+VIDEO

2. What is your gender?
   - MALE
   - FEMALE

3. How old are you? _____________________________

4. What is your first language? _____________________________

5. Have you learned a foreign language other than English?
   - YES (If yes, which language is it? ____________ years)
   - NO

6. Have you ever lived in an English-speaking country?
   - YES (If yes, how long? ____________ years)
   - NO

7. Did you go to school in the U.S. or any English-speaking country?
   - YES (If yes, how long? ____________ years)
   - NO

8. What type of school did you attend? (Only if yes to #7) Check all applicable boxes.
   - ELEMENTARY SCHOOL
   - MIDDLE SCHOOL
   - HIGH SCHOOL
   - LANGUAGE PROGRAM
   - COLLEGE/UNIVERSITY
   - GRADUATE SCHOOL
   - OTHER (SPECIFY: ____________________________)

9. At what age did you start learning English? AT AGE ______________

10. How were you first exposed to English?
    - CLASS
    - MEDIA (TV, RADIO, ETC.)
    - COMMUNICATION WITH NATIVE SPEAKERS
    - STUDY ABROAD
    - OTHER (SPECIFY: ____________________________)

11. How long have you studied English? _____________________________

12. How would you rate your English proficiency?
    - NATIVE
    - ALMOST NATIVE
    - GOOD
    - FAIR
    - POOR
13. How would you rate the instruction?
   - EXCELLENT
   - VERY GOOD
   - GOOD
   - FAIR
   - POOR

14. How useful was the instruction in remembering words that were taught?
   - VERY
   - FAIRLY
   - MODESTLY
   - NOT AT ALL
   - DON’T KNOW

Thank you very much for your participation!
ENGLISH LANGUAGE INSTITUTE

LISTENING COMPREHENSION TEST

This is a test of your ability to understand spoken English. You will hear the voice either ask a question or make a statement. To show that you have understood what was said, you are to select the ONE answer choice you think is correct, and circle that choice on your test.

Here is an example. Listen carefully to the question, then choose one of the answers given below.

Example A: a. I am.
    ⑤ Tomorrow
    c. At home.

The correct answer is b. Tomorrow. Choice b. has been marked to show that b. is the correct answer.

Now here is an example of the statement type of problem.

Example B: ③ The camera is expensive.
    b. The desk is expensive.
    c. The camera and the desk are expensive.

The correct answer is a. The camera is expensive. Choice a. has been marked to show that a. is correct.

Please be quiet and listen carefully. None of the questions or statements can be repeated. Now, open the test booklet to problem number one.
1. a. Yes, it is.  
   b. Yes, it was.  
   c. Yes, I have.  

2. a. No, she doesn’t.  
   b. Mary.  
   c. No, she hasn’t.  

3. a. Mother did the dishes.  
   b. Father did the dishes.  
   c. Gail did the dishes.  

4. a. Neither of them will sing.  
   b. One of them will sing.  
   c. Both of them will sing.  

5. a. Yes, we are.  
   b. Yes, he is.  
   c. Yes, they are.  

6. a. We went to Europe and Japan.  
   b. We went only to Europe.  
   c. We went only to Japan.  

7. a. For about five minutes.  
   b. To thank her.  
   c. In five minutes.  

8. a. Neither is good.  
   b. Both are good.  
   c. Only his is good.  

9. a. My brother is.  
   b. My brother does.  
   c. To my brother.  

10. a. There has been no snow in the past.  
    b. There has been less snow in the past.  
    c. There has been more snow in the past.  

11. a. Susie and Jane listened to the story.  
    b. Jane wrote the story.  
    c. Susie and Jane wrote the story.  

12. a. Yes, it is.  
    b. Yes, it has.  
    c. Yes, she has.  

    b. Sally thought Jane knew him.  
    c. Neither Sally nor Jane knows him.  

14. a. No, I begin there next year.  
    b. No, I’ve finished already.  
    c. No, I finish there next year.  

15. a. He didn’t think Alice would come.  
    b. He thought Alice would come.  
    c. He doesn’t think Alice will come.  

16. a. Yes, I have.  
    b. Yes, it is.  
    c. Yes, it was.  

17. a. No, he doesn’t.  
    b. My brother.  
    c. No, he hasn’t.  

18. a. We went to Florida.  
    b. We went to California.  
    c. We went to Canada.  

19. a. Both of them eat it.  
    b. Neither of them eats it.  
    c. One of them eats it.  

20. a. Yes, he is.  
    b. Yes, they are.  
    c. Yes, it is.  

21. a. We bought only the motorcycle.  
    b. We bought only a car.  
    c. We bought a car and a motorcycle.  

22. a. In a few days.  
    b. For about a week.  
    c. To see my parents.  

23. a. He likes only one.  
    b. He likes neither.  
    c. He likes both.  

24. a. His sister is.  
    b. His sister did.  
    c. To his sister.  

128
25. a. Usually more people come.
   b. Usually no people come.
   c. Usually fewer people come.
26. a. Sam and Mary heard it.
   b. Sam wrote it.
   c. Sam and Mary wrote it.
27. a. Yes, it has.
   b. Yes, she has.
   c. Yes, it is.
28. a. Mrs. Jones thought Mr. Jones knew them.
   b. Neither Mr. nor Mrs. Jones knew them.
   c. Mrs. Jones knew them.
29. a. No, I sold it.
   b. No, I will get one.
   c. No, I want to sell it.
30. a. John didn’t want the friend to come.
   b. John wants the friend to come.
   c. John wanted the friend to come.
31. a. Yes, it was.
   b. Yes, I have.
   c. Yes, it is.
32. a. No, he doesn’t
   b. Marc.
   c. No, he hasn’t.
33. a. His favorite class is English.
   b. His favorite class is history.
   c. His favorite class is math.
34. a. Both boys like them.
   b. Neither boy likes them.
   c. One boy likes them.
35. a. Yes, they did.
   b. Yes, it did.
   c. Yes, he did.
36. a. We went skiing and skating
   b. We only went skiing
   c. We only went skating
37. a. Her birthday.
   b. In two hours.
   c. For about two hours.
38. a. Only one is windy.
   b. Neither is windy.
   c. Both are windy.
39. a. To Mike.
   b. Mike’s.
   c. Mike did.
40. a. She never looked as angry.
   b. She has looked more angry.
   c. She never looks angry.
41. a. Dave and John saw it.
   b. John wrote the article.
   c. Dave and John wrote the article.
42. a. No, it hasn’t.
   b. No, he hasn’t.
   c. No, it isn’t.
43. a. Neither Bill nor Jack knows him.
   b. Bill knew him.
   c. Bill thought Jack knew him.
44. a. No, I finished it.
   b. No, I’m reading it now.
   c. No, I haven’t seen it.
45. a. Bill wanted her to hire one.
   b. Bill doesn’t care if she hires one.
   c. Bill didn’t want her to hire one.
MICHIGAN QUESTIONS (Vers. 4)

Examples:

Example A: When are you going?
Example B: The camera on the desk is expensive.

Questions:

1. Was that a good movie you saw?
2. Does Mary know the assignment already?
3. Gail did the dishes because father was too tired and mother had to go out.
4. Laura won't sing and Cathy won't either.
5. Do you know if Grandfather is coming for dinner on Thursday?
6. We planned to vacation in Europe, but went to Japan, too.
7. What do you want to see her for?
8. Her job isn't very good, but her husband's is.
9. Who's talking to Mr. Smith?
10. I've never seen snow here.
11. I read Susie and Jane's story.
12. Has mother's letter arrived yet?
13. Sally told Jane who the new boy was.
14. Haven't you left the university yet?
15. Charles would be surprised if Alice came to his party.
16. Have you taken this train before?
17. Has your brother seen the movie?
18. We went to California for our vacation because Florida's too hot and Canada's too cold.
20. Can you see if the people are in their seats?
21. We wanted to buy a motorcycle, but got a car instead.
22. How soon are you going to New York?
23. He doesn't like tea, nor does he like coffee.
24. Who was Bill talking to just now?
25. We've seen more people in this park.
26. I played Mary Sam's composition.
27. Is Sally's dinner invitation here yet?
28. Mr. and Mrs. Jones wondered who their new neighbors were.
29. Have you gotten rid of your bicycle yet?
30. John was happy that Dave didn’t bring his friend.
31. Is it snowing right now?
32. Who went to Europe this summer?
33. Gary likes his English class the best, because math is too hard and history is too dull.
34. Alan doesn’t like girls, but Jim does.
35. Do you know if the car his grandparents ordered arrived yet?
36. We intended to ski and skate, but didn’t have time to skate.
37. How long will the party last?
38. Chicago is pretty windy, but so is Detroit.
39. Who brought these books to you today?
40. John’s never seen his wife looking more angry.
41. I showed Dave and John my article.
42. Has father received his coat from the cleaners?
43. Bill asked Jack who their new teacher was.
44. Are you still reading that book?
45. Bill was upset that his wife hired a cook.
XI. References


effects on L2 vocabulary acquisition in CALL literature. ReCALL, 23(2), 135-159.

on video comprehension and incidental vocabulary learning. Language Learning &
Technology 18(1), 118-141.

auditory and visual presentation modes. Journal of Educational Psychology, 87(2),
319.

PD Pearson, and R. Barr (Eds.), Handbook of reading research, Vol. 3 (pp. 269-
284).

RELJ Journal, 13(1), 14-36.


http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-
Test-information-and-specifications.pdf

Cambridge University Press.

31(7), 9-13.

Teaching, 97, 110.


