5-19-2017 12:00 AM

Short and long-term vocabulary learning and retention through multimedia glossing: A mixed methods research

Nasrin Ramezanali, *The University of Western Ontario*

Supervisor: Dr. Farahnaz Faez, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Education

Abstract

Vocabulary development is significant for mastering a second/foreign language. There are several approaches for vocabulary instruction, including glossing. Glossing is a context-based technique that represents information on target words via definitions, explanations, synonyms, pictures, sounds, and videos. Glossing can also be used electronically through texts, pictures, audios, and video/animations. Studies on different gloss combinations in multimedia learning environments have led investigators to inconclusive findings. The present study examined which mode(s) of gloss presentation (L2 definition, aural, and video/animation) is effective for learners' short and long-term vocabulary learning and retention.

Utilizing a mixed methods approach, 132 intermediate language learners formed one control and three experimental groups. The experimental groups received target words in different glossing modes; the control group received no glossing instruction. ANCOVA and paired samples t-test were used to analyze the pre/post-test data. Learners' attitudes and perceptions towards glossing modes were also examined through a questionnaire and interviews.

The results showed that glossing was significantly more effective than non-glossing strategy for participants' short-term retention in both productive recall and multiple-choice productive recognition tests; and partially effective for their long-term retention. Additionally, in both vocabulary measurements, L2 definition and video/animation glossing as well as L2 definition and audio glossing were more effective than L2 definition alone for most test sessions; but since L2 definition alone was also effective for few test sessions, the findings cannot be generalized largely. The results of the questionnaire and interviews showed that the participants preferred L2 definition and video/animation glossing over the two other modes.

Keywords: Glossing, Vocabulary learning, Short and long-term word retention, Multimedia learning.

Dedication

To My Beloved Husband, Alireza

Acknowledgements

Without the support, encouragement, and contributions from my professors, mentors, colleagues, friends, and family, the successful completion of this PhD work would not have been possible. I take this opportunity to appreciate them for their expert advice and wisdom, invaluable comments, and generous time. They enriched my knowledge and made my graduate journey a memorable chapter in my life.

First and foremost, I am grateful to my thesis supervisor, Dr. Farahnaz Faez, for her amazing supervision. Her outstanding knowledge and professional teaching-related expertise inspired me greatly during my academic life at UWO, and she extended new horizons for me in the field. I appreciate her helpful guidelines and suggestions, unfailing support, great patience, and strong management skills in helping me resolve the obstacles during my PhD life. I feel fortunate to have had her support for four years.

I would like to extend my heartfelt gratitude to my thesis supervisory committee, Professor Stuart Webb and Dr. Steve Bird, for their compassionate, inspirational and thoughtful guidance, careful reading, and insightful comments. It has been a great honor and privilege to work with Professor Webb and learn from his invaluable and constructive feedback throughout my dissertation. Dr. Bird also inspired me with his knowledge and expertise on the quantitative component of this research. I am thankful to both for their encouragement and kind words.

I am also grateful to my external examiner, Dr. David Wood, to my faculty examiners, Dr. Suzanne Majhanovich and Dr. Julie Byrd Clark, and to my university examiner, Dr. Nadine de Moras, for their insightful comments and questions during my oral defense. I would like to extend an appreciation to my professors at UWO with their invaluable contributions in various components of the program.

I would like to extend my sincere thanks to Dr. Mehdi Delrobaei for his support and devotion of time on designing the multimedia instruments of this study on the website. Also, my warmest gratitude goes to Dr. Masood Siyyari for his great expertise in SPSS software, and his absolute dedication of time, support, and thorough analyses of the quantitative component of this study along with the detailed explanations.

This study was made possible by the presence of the passionate students at a language institute in Iran who devoted their time participating in this research. I also appreciate Ms. Yalabadi for her amazing assistance, support, and management during the data collection phase of this study.

I thank all my colleagues and friends who offered collegial guidance over the years for my educational growth. I am grateful for the financial support provided by UWO, and the research office at the Faculty of Education for various conference scholarships that allowed me to present my PhD work at several local and international conferences. My deepest thanks go to my PhD cohort, particularly Carolyne and Joelle, for all their time and unfailing support during my academic life at UWO. Carolyne has always been a great friend and nice fellow since I stepped into the program. Words cannot express her kind and stimulating encouragement as well as valuable feedback and technological expertise on earlier drafts of my work. Joelle also provided me with strong and steady support in numerous ways. Her lovely voice on audio instruments of this research is only one among many of her dedications to my work.

I am also grateful to all the support and encouragement I received from my lovely parents and sisters throughout my PhD education. It feels delightful to know that accomplishing this PhD has made them proud.

Last, but in no way least, I want to thank my beloved husband, Alireza, for his sincere love, concern, and support. Without him in my life and his whole-hearted encouragement, the endeavour of this work would not have been possible. I owe him a lot, now and forever.

Table of Contents

List of Figures

List of Tables

List of Appendices

CHAPTER ONE: INTRODUCTION

Background and Significance of the Study

Vocabulary development is one of the most important aspects of foreign language (FL) or second language (L2) learning and teaching (Hunt & Beglar, 2005; Knight, 1994). Acquisition of words is also the basis for communication in FL/L2 contexts and “an essential part of mastering a L2” (Schmitt, 2008, p. 329). In this regard, Wilkins (as cited in Milton, 2009) notes that “without vocabulary, nothing can be conveyed” (p. 111). As such, Laufer (1997) asserts "no text comprehension is possible, either in one's native language or in a foreign language without understanding the text's vocabulary" (p. 20). Thus, developing a rich vocabulary is both a priority and a challenge for FL/L2 learners. When FL/L2 learners are faced with a reading passage, their lack of vocabulary knowledge, as a linguistic constraint (Rassaei, 2017; Yusuf, Sim, & Su’ad, 2014) is their major obstacle to the comprehension of the reading text. If that text has too many new words, FL/L2 learners quickly become discouraged (Grabe & Stoller, 1997), and refuse to continue reading the passage. Moreover, FL/L2 learners and/or teachers are familiar with the disconcerting experience of trying to recall, without success, a word which has only been recently encountered and used, or a word, which has been in their vocabulary for a long time, but seems to elude them when it is needed. Therefore, in order to deal with these challenges, FL/L2 learners need multiple exposure to L2 vocabulary in various contexts through a variety of vocabulary instruction techniques and strategies (Nation, 2011; Schmitt, 2008). These vocabulary techniques can assist FL/L2 learners to cope with unfamiliar words (Harley, 1995), recall them promptly over long term, and apply the acquired words in communicative contexts correctly (Nagy, 2005; Read, 2004).

The traditional mediums of vocabulary teaching include word-lists, dictionary use, workbooks, teacher-materials, using vocabulary cards, and negotiating vocabulary meanings (Hunt & Beglar, 2005; Hulstijn, 1996), and glosses— "a brief definition or synonym, either in L1[1] or L2, which is provided with the text" (Nation, 2013, p. 238). Although consulting a dictionary, as an example of one vocabulary aiding strategy, is the easiest way to fulfil the need of L2 learners, it lowers their reading speed and may confuse them with several meanings (Luppesau & Day, 1993); thus, sometimes leading the learners to wrong inferences or guesses (Hulstijn, 1992). However, innovative and technologically-based techniques are emerging that facilitate FL/L2 learners' vocabulary learning. In other words, there are new instructional and context-based tools available to help FL/L2 learners to learn new words and retain them longer. As such, there is a need to examine effective and practical FL/L2 pedagogical techniques for vocabulary instruction that encourage FL/L2 learners to process the meanings of the words and enhance their long-term recollection of vocabulary (Al-Seghayer, 2003). One such instructional and context-based vocabulary technique is glossing in computerized and multimedia fashion, which substitutes for the traditional mediums of vocabulary learning, such as dictionary use (Yanguas, 2009). It also has the consequent effects of saving FL/L2 learners' time and effort (Jacobs, 1994; Lomicka, 1998; Roby, 1991) and increasing the comprehension of the text (Hulstijn, 1992; Jacobs, 1994; Leffa, 1992; Watanabe, 1997). Additionally, glossing has the potential for vocabulary learning

[1] First Language

(Webb, 2010); and circumvent/avoid the need for consulting a dictionary. The new medium thus provides a dynamic L2 vocabulary learning environment in which ample vocabulary learning materials from multiple sources are implemented.

Knowing that vocabulary is an essential building block of language (Schmitt, Schmitt & Clapham, 2001); and that vocabulary knowledge has a critical role in improving communication skills (Zarei & Mahmoodzade, 2014), enhancing reading comprehension (Zandieh & Jafarigohar, 2012; Nation, 2001; Watanabe, 1997), and acquiring a new language (Nation, 2001), the present doctoral research aimed to examine the effectiveness of multimedia glossing in helping FL/L2 learners to increase their vocabulary knowledge and foster their short and long-term word retention. This study is important as it has provided insight crucial to vocabulary learning and retention. Findings from this study are of particular interest to educators, researchers, material developers, and syllabus designers who are looking for appropriate ways to increase learners' vocabulary knowledge, recollection, and comprehension.

Definition of Conceptual Terms / Key Terms

The following operational definitions were used in this study:

Glossing

Glossing or annotation is information on new words conveyed via definitions, marginal explanations of a key word in L1 or L2, synonyms, examples, translations, adequate context, pictures, audios and videos (Chen, 2016; Jung, 2016; Marefat, Rezaee, & Naserieh, 2016; Nation, 2001; Stewart & Cross, 1993, 1991; Samian, Foo, & Mohebbi, 2016; Webb, 2010). Khezrlou and Ellis (2017) define gloss as the definitions or translations of the new words in the passage that aid learners' comprehension. Glossing is one practical vocabulary instruction technique that can take several forms (verbal, visual & aural). Glossing has been shown to be effective for promoting FL/L2 vocabulary learning (Al-Seghayer, 2001; Cheng & Good, 2009; Choi, 2016; Chun & Plass, 1996; Hong, 2010; Hulstijn, Hollander, & Greidanus, 1996; Jacobs, Du Fon, & Hong, 1994; Jung, 2016; Ko, 2012; Türk & Erçetin, 2014; Watanabe, 1997; Zandieh & Jafarigouhar, 2012), and increasing word coverage[2] (Webb, 2010).

Gloss types in the present study.

This study adopted the verbal (L2 definition of the new words), aural (audio recording of the pronunciation of the new word as well as its definition), and visual (video/animation of the new words along with its definition) as the three forms of glossing.

Multimedia, Multimedia Instruction and Multimedia Learning

Mayer (2014) refers to multimedia as "presenting words (such as printed text or spoken text), and pictures (such as illustrations, photos, animation, or video)" (p. 2). Multimedia instruction is thus described as any attempt to present words and pictures that foster learning (Mayer & Moreno, 2003); and multimedia learning is simply learning from words and pictures (Mayer & Moreno, 2003).

Multimedia/electronic glossing[3].

The impact of glossing individual vocabulary via electronic modes and media, known as multimedia/electronic glossing, is one type of vocabulary learning environment that has recently received increasing attention (Türk &

[2] Word coverage is defined as the percentage of known words in a context (Webb, 2010).
[3] Multimedia and electronic glossing are used interchangeably in this dissertation.

Erçetin, 2014; Yanguas, 2009). Multimedia glossing emerged as a result of the integration of computer and multimedia technology (known as computer assisted language learning or CALL) with glossing in the domain of vocabulary instruction. Multimedia glossing takes the form of not only text/words but also of pictures and graphics, audios, icons, videos, and animations. This doctoral research followed Mayer's (2014) definition of multimedia. Accordingly, I referred to pictures as materials in "*pictorial form*, such as […] dynamic graphics, including animation or video"; and words as "materials in verbal form, such as printed text or spoken text" (Mayer, 2014, p. 2). Multimedia/electronic glossing in this research context is thus defined as the application of gloss in combination with text, audio, and video/animation in a multimedia-based learning environment, such as the language classroom where L2 learners sat in front of a computer monitor. Also, *text* was represented as L2 definition of a new word.

Simultaneous multimedia glossing. According to one of the principles of multimedia learning, the temporal contiguity principle[4], multimedia information such as texts, pictures, audios, and or videos/animations can be presented either simultaneously (where the materials are displayed at the same time), or successively/interactively (where the corresponding materials are separated in time) (Mayer & Fiorella, 2014; Moreno & Mayer, 2002a, 2002b). Thus, glossing through multimedia information can be either displayed concurrently or one after the other.

[4] See research rationale in this chapter for the discussion of the temporal contiguity principle.

Word Retention/Recollection

Retention/recollection is defined as the ability to provide the meaning of a new word after a given period of time. There are two types of word retention: short-term retention and long-term retention.

Short and long-term word retention.

This study has taken the immediate performance of L2 learners' vocabulary tests after instruction as the short-term word retention; and the time span of two weeks/14 days after the intervention/instruction as the long-term word retention.

Target Glossed Words (TGWs)

The new words that were subject to glossing during the intervention/instruction were referred to as the target glossed words in this study. TGWs were tested both after instruction and two weeks later to evaluate the amount of lexical knowledge gained.

Statement of the Problem/Gaps in the Literature

Researchers in multimedia glossing seek to know whether, if at all, and how types of media will facilitate FL/L2 learners' acquisition of new vocabulary (Akbulut, 2007; Salem, 2006). Despite the fact that several studies have been conducted on the impact of multimedia/electronic glossing on FL/L2 vocabulary learning (Akbulut, 2007; Al-Ghafli, 2011; Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b; De Ridder, 2002; Lyman-Hager & Davis, 1996; Tabatabaei & Shams, 2011; Türk & Erçetin, 2014; Yanguas, 2009; Yoshii & Flaitz, 2002; Yun, 2011), and word retention (Salem & Aust, 2007; Salem, 2006; Kost, Foss & Lenzini, 1999; Zandieh & Jafarigouhar, 2012), the findings still show inconclusive, insufficient, and controversial evidence as to what gloss combination is more effective in facilitating vocabulary learning and enhancing long-term word recollection. As an example, to name just a few, Chun and Plass (1996a) and Plass, Chun, Mayer, and Leutner (1998) found that visual glossing of text and picture was more effective than visual glossing of text and video for L2 vocabulary learning, whereas Al-Seghayer (2001) and Lin and Tseng (2012) showed that integrating text and video led to better vocabulary learning and retention compared to their integration with text and picture. The findings of Akbulut (2007) revealed an equal and positive effectiveness or "zero effectiveness" (Mohsen & Balakumar, 2011, p. 151) for the two modalities of text and pictures and text and videos in facilitating vocabulary learning and word retention compared to text definition alone due to some factors such as learners' cultural and linguistic backgrounds, differences in their target language and proficiency level, as well as level of word difficulty and assessment types (Mohsen & Balakumar, 2011). Similarly, in my master's thesis research (Ramezanali, 2003) on the effectiveness of different annotation types on L2 vocabulary learning, I found that the three presentation modes of text, pictures, and videos had similar effects on vocabulary learning of EFL learners, and that textual definition alone was shown to be more effective for word learning than the other two modes. Thus, further research is required to examine which gloss type(s) better aid FL/L2 learners in improving L2 vocabulary learning and enhancing short and long-term word retention. Likewise, Boers, Warren, He, and Deconinck (2017) found that between the two glossing modes of textual and pictorial, the latter one did not help learners to learn and retain words any better than providing the learners with textual explanations.

Some researchers have investigated different combinations of gloss types in multimedia-learning environments that involve mainly texts, pictures, and videos (Akbulut, 2007; Lin & Tseng, 2012; Plass et al., 1998; Salem, 2006), texts, pictures, and audios (Salem & Aust, 2007; Salem, 2006; Yeh & Wang, 2003), pictures, audios, and videos (Al-Ghafli, 2011; Sakar & Erçetin, 2005), and texts and pictures (Tabatabaei & Shams, 2011; Yanguas, 2009; Yoshii & Flaitz, 2002). However, very little is known about the impact of glossing in combination with texts, audios, and videos/animations on L2 vocabulary learning and long-term word retention in a computerized-learning environment. For the studies that incorporated the audios in the gloss combination, findings show either no significant difference (Yeh & Wang, 2003) or a distracting effect (Kim & Gilman, 2008) on the addition of sound to the glossed words for L2 vocabulary learning. Therefore, more research is needed to consider the effect of multimedia/electronic glossing on L2 vocabulary learning and word retrieval with the addition of audio glossing.

Moreover, research has shown that presenting information through simultaneous display condition helps FL/L2 learners to build mental connections between the information and the presentation mode (Mayer & Moreno, 2003; Mayer, 1997). The simultaneous exposure also results in better transfer of information and retention (Mayer & Sims, 1994; Mayer & Anderson, 1991). Yet, only a few studies on the domain of multimedia glossing have investigated the effects of simultaneous versus interactive/successive display mode of multimedia information on vocabulary learning (Türk & Erçetin, 2014). In their study, Türk & Erçetin (2014) considered interactive glossing as allowing FL/L2 learners to have control over the selection of gloss type (text

& picture). However, there is a need to conduct more research on the domain of multimedia glossing that only takes into account the simultaneous gloss presentation of text, audio, and video/animation on L2 vocabulary learning in terms of short and long-term word retention.

Finally, very few studies have used a mixed methods research design to determine the impact of multimedia gloss types on long-term word retention (Akbulut, 2007; Ko, 2005; Yanguas, 2009). Some studies have qualitatively investigated the effect of electronic glosses on reading comprehension and word retrieval (Lenders, 2008; Lomicka, 1998), and others have conducted quantitative research to examine the applicability of gloss types on word knowledge, reading fluency, and vocabulary retention (Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b).

Research Purpose and Scope of the Study

To address these methodological concerns, the purpose of this study was to examine the effectiveness of meaningful[5], contextual, and multimedia learning tools of simultaneous textual[6], aural, and video/animation glossing to improve L2 learners' vocabulary learning. The secondary goal of the study was to investigate if simultaneous multimedia glossing versus single mode glossing fostered L2 learners' short and long-term word retention.

This study was framed within the scope of vocabulary acquisition/learning of L2 learners over short and long-term through implementing both dual and single glossing

[5] See research rationale (simultaneous display of glossing) for the discussion on Meaningful learning.
[6] Textual refers to L2 definition of a word in English.

modes. The two gloss combinations of simultaneous L2 definition and audio glossing as well as simultaneous L2 definition and video/animation glossing functioned as the dual glossing types; and L2 definition alone accounted for the single glossing mode. I sought to investigate if these glossing modes increased learners' vocabulary learning and influenced their short and long-term word retention. The study took place with group of L2 adult learners across one level of language proficiency (intermediate) in an EFL context.

Research Rationale

This section outlines my justifications for choosing different components of the study that included: (a) L2 learners' intermediate level of language proficiency; (b) gloss combination of L2 definition, audio, and video/animation; and (c) simultaneous display mode of multimedia glossing. I have also indicated my personal rationale to pursue this doctoral research.

Intermediate Language Proficiency Level

I rationalized the selection of intermediate L2 learners according to cognitive load theory (Chandler & Sweller, 1991), which posits that learning of any type happens when learners are cognitively capable of dealing with the presented materials. One of the assumptions of cognitive load theory is the limited capacity of human working memory that restricts learners to process new information. L2 learners might have challenges for processing the glosses with different modalities such as textual, aural, and visual glossing depending on their language proficiency level. In the case of vocabulary learning via multimodal tools of vocabulary glossing, the findings of Plass and colleagues (1998) showed that language learners with low abilities in verbal and visual communication skills may not fully experience the benefits of multimedia glosses due to the high cognitive load of the materials when presented with two or more types of gloss annotations for vocabulary learning. The results further demonstrated that learners with low visual and verbal proficiencies performed poorly on vocabulary tests when they received visual annotations (pictures & videos). Thus, learners with less language proficiency may encounter some new target items beyond the development stage of language learning; and consequently, may not be able to utilize all the benefits of multimedia glosses while learning new words due to the high cognitive load of the tools (Loewen & Erlam, 2006; Sweller, 1994; Vahedi, Ghonsooly, & Pishghadam, 2016). Similarly, other findings reveal that L2 learners at intermediate and higher than average proficiency levels benefited more from vocabulary glossing, because they possessed sufficient L2 competence and knew how to use and process the glosses effectively (Abraham, 2008; Erçetin, 2003; Jacobs et al., 1994; Ko, 2005). In other words, glossing has a "large positive effect over time" for intermediate-level learners (Abraham, 2008, p. 210). Therefore, since proficiency level was a determining factor in gloss effects, and it was unlikely that all levels of L2 learners benefited equally from gloss exposure (Cheng & Good, 2009), L2 learners with an intermediate level of proficiency were selected for this study.

Gloss Combination

The rationale for choosing different gloss combinations was multi-fold. First, by using different gloss combinations, I aimed to expose L2 learners to a rich vocabulary learning environment. Second, some researchers have acknowledged that due to the individual differences of L2 learners for vocabulary learning, presenting information via a variety of gloss modes (textual, aural, & visual) can be tailored to suit their differences, and will provide them with more opportunities to learn new words in their preferred modes (Al-Seghayer, 2003; Plass et al.,1998). Thus, the participants of this study were exposed to different glossing modes to cope with their individual differences. Furthermore, research has shown that glossing alone has limited effect on long-term vocabulary retention (Jacobs et al., 1994; Watanabe, 1997). Therefore, in order to tackle this problem, various gloss combinations were used to mentally engage learners to recall and retain the learned words. Finally, I based the selection of different gloss combination upon Wittrock's (1974) generative theory of learning. Wittrock suggested that learning is facilitated when information is presented in multiple forms rather than a single mode. In this way, learners can select relevant information to construct meaning. Consequently, different gloss combinations were practiced in this research to allow L2 learners to rebuild meaning associations while learning target vocabulary with different glossing modes.

The reason why the combination of L2 definition and audio glossing versus L2 definition and video/animation glossing was used in a multimedia-based learning environment was also attributed to the two principles in multimedia learning: (a) multimedia principle, and (b) voice principle. These two techniques engage L2 learners in vocabulary learning. According to Mayer (2014), in voice principle, "human voice for spoken words" is used (p. 63), which accounted for the audio glossing in this study; and, in multimedia principle, learners can take advantage of words and pictures (either static or dynamic) rather than words alone, which accounted for the L2 definition and video/animation glossing.

Simultaneous Display of Multimedia Glossing

My reasoning to display multimedia glosses simultaneously stemmed from one of the multimedia instruction principles, namely the temporal contiguity principle (Mayer & Fiorella, 2014; Moreno & Mayer, 2002a, 2002b). Moreno and Mayer (2002b) contend that learners "learn more deeply when corresponding portions of the narration and animation are presented at the same time than when they are separated in time" (p. 95). The major theoretical rationale for simultaneous presentation is based on the cognitive theory of multimedia learning and its assumptions that meaningful learning happens when learners are able to build mental connections between verbal and visual representations in their working memory (Mayer, 2014; Mayer, 2005; Moreno & Mayer, 2002a), and that the corresponding materials are presented simultaneously rather than successively for better learning (Mayer & Fiorella, 2014; Mayer, 2008; Rusanganwa, 2015). In other words, L2 learners can hold both verbal and visual information in working memory when narration (printed & spoken words) and images (pictures & video/animations) are coordinated in time (Mayer, 2008; Moreno & Mayer, 2002a). Thus, this multimedia presentation reduces cognitive load (Mayer, Moreno, Boire, & Vagg, 1999; Mayer & Moreno, 2003; Mayer, 1997). According to Moreno and Mayer (2002 a), "multimedia instructions should be designed in ways that minimize the chances of overloading the learners' cognitive system" (p. 108). Therefore, the present research study offered an engaging vocabulary learning environment where the participants were presented with verbal (textual & aural), and visual (textual & video/animation) information simultaneously rather than successively in order to interact with the reading passage effectively. The interaction may lead them to a better vocabulary gain and performance (Türk & Erçetin, 2014).

Personal Rationale

I chose the domain of vocabulary learning as the focus of this doctoral research because I had always struggled to find the most efficient strategies that would assist me in retaining new words longer and improve my reading comprehension and effective communication when learning English. I noticed that when I learned unfamiliar words with their pronunciation/audio transcriptions, or animation displays, I could grasp them better and retrieve them faster. Hence, I believed that looking for effective multimedia-based pedagogical vocabulary learning strategies, which could assist L2 learners in learning words and remembering them longer, would make for a unique study and would add to the limited literature currently available in this area.

Research Questions

As stated earlier, the major goal of this research was to investigate the effectiveness of simultaneous multimedia glossing of textual, aural, and video/animation on L2 learners' vocabulary learning using a mixed methods research methodology. A secondary goal was to examine if simultaneous multimedia glossing versus single mode glossing fostered L2 learners' short and long-term word retention. To address these concerns, this study was guided by the following research questions and sub-questions:

1. Do different glossing modes (L2 definition alone, L2 definition & audio glossing, or L2 definition & video/animation glossing) have any significant impact on L2 learners' vocabulary learning and short and long-term word retention?

1.1. Does glossing have any significant impact on L2 learners' vocabulary learning for short-term word retention?

1.2. Does glossing have any significant impact on L2 learners' vocabulary learning for long-term word retention?

2. Do different glossing modes (L2 definition alone, L2 definition & audio glossing, or L2 definition & video/animation glossing) have any significant impact on L2 learners' vocabulary learning and short and long-term word retention?

2.1. Which glossing mode (s) contributes significantly to L2 learners' short-term word learning and retention?

2.2. Which glossing mode (s) contributes significantly to L2 learners' long-term word learning and retention?

3. What are L2 learners' attitudes and perceptions towards simultaneous multimedia glossing? Which glossing mode (s) do they prefer, and why?

Overview of Dissertation Chapters

This doctoral dissertation is divided into six chapters. Chapter *one* provides some preliminary background research regarding vocabulary learning via different glossing modes. It presents the purpose and significance of the study, and rationalized the choice of gloss combination, language proficiency level, and my personal motive to pursue this project. Chapter *two* explores the two underlying theoretical constructs that framed this study: dual-coding theory (Paivio, 2007, 1986) and cognitive theory of multimedia learning (Mayer, 2014, 2005). It also explores the relevant literature regarding glossing and multimedia vocabulary learning. Chapter *three* explains the methodological paradigm as well as the research methods used to collect the data. The findings and analyses are presented in chapter *four*, followed by a thorough discussion of the results in relation to the literature (chapter *five*). Finally, chapter *six* presents some concluding remarks and a summary of the findings in relation to each research questions. The limitations to this research are stated; and the areas for future research are then suggested.

Overview of the Chapter

In this chapter, I introduced the background and significance of the study, followed by the operationalization of the conceptual terms. I then discussed the statement of the problem, the purpose and scope of this study as well as the rationales. The research questions were then presented along with the null hypotheses. The chapter ended with the outline of this dissertation.

CHAPTER TWO: THEORETICAL FRAMEWORK AND LITERATURE REVIEW

Introduction

The purpose of this study was to examine the impact of multimedia glossing on L2 learners' vocabulary learning for short and long-term word retention. Several studies have reported that the application of multimedia learning[7] in language education, in general, and word acquisition, in particular, has resulted in building meaningful connections between words (spoken or printed) and pictures (static or dynamic), and consequently maximized vocabulary learning (Mayer, 2014, 2009; Plass et al., 1998). This dissertation draws on two prominent theories that explain the value and effect of multimedia presentations in language learning environments: (a) Paivio's (2007, 1986) dual coding theory, and (b) Mayer's (2014, 2005) cognitive theory of multimedia learning. These theories complemented the goals of this study and provided strong rationale for vocabulary learning in multiple modes of annotations. Research has shown that vocabulary glossing through different annotation modes, media, and forms improves L2 word acquisition "when presented digitally on a computer screen" (Al-Seghayer, 2016, p. 179). Thus, it is necessary to provide language learners with multimedia-based learning tools that help them to increase the word knowledge in their vocabulary reservoir and enhance their short and long-term word recollection.

[7] Review multimedia learning in chapter one.

This chapter is divided into two parts. Part one presents the underlying features of the two theories and discusses the implications for multimedia learning, in general, and multimedia glossing, in particular. Part two explores the current literature in the field regarding multimedia glossing, and relates it to the theoretical framework of the study as well as the findings.

Part One: Theoretical Framework

The Dual-Coding Theory (DCT)

The dual coding theory (DCT) (Paivio, 1971, 1986) is a cognitive theory explaining the powerful effects of mental imagery on the mind and memory. DCT is based on the idea that the formation of mental images aids in learning (Reed, 2010). The theory postulates the existence of two separate, but interrelated coding systems, which process and store information in the memory: (a) a verbal system and (b) a non-verbal/visual system. The verbal system stores linguistic information/units (such as text & sound) in sequential units called "logogens." The non-verbal/visual system processes visual information/units (such as pictures, animations, &/or videos) and keeps them in units called "imagens." The two systems are linked together through referential connections.

The main assumption of DCT is that verbal and non-verbal/visual modes are processed by two different coding systems; however, these verbal and non-verbal/visual systems interact, and the activation of both systems results in better recall (Al-Seghayer, 2001; Paivio, 1991). According to Paivio (1986) and Clark and Paivio (1991), there are three different processing levels that take place within or between verbal and non-verbal/visual systems: representational, referential, and associative processing.

Representational processing refers to the activation of the verbal and visual representation by a stimulus, in that words activate verbal representation, and pictures activate visual representation. *Referential* processing refers to the activation of either system by the other one, meaning the images or objects activate words, and words activate images or objects. *Associative* processing refers to the activation of additional information in representational or referential systems. At this level, the associative connections between words and sounds (i.e., linguistic units) in the verbal system and images in the visual system are activated. Figure 1 illustrates how verbal and non-verbal/visual systems are linked together through referential connections (Paivio, 1986).

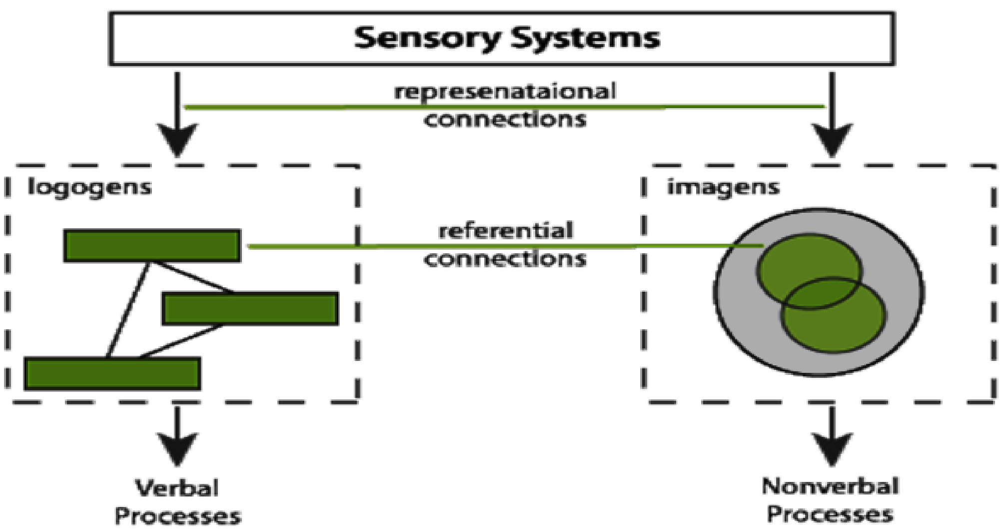


Figure 2. 1 Paivio's Dual-Coding Theory (1986, p. 67)

Research has shown that words which are associated with actual objects or images are acquired more easily and retained more effectively than those which are presented alone (Clark & Paivio, 1991). Since the purpose of this study was to examine L2 learners' vocabulary learning in terms of both short and long-term word retention by exposing them to different combinations of verbal representations (i.e., L2 definition & audio glossing) and visual representations (i.e., L2 definition & video/animation glossing) in a multimedia-based learning environment, dual-coding theory offered a solid and base framework for developing the verbal and visual aids. Thus, it was assumed that when L2 learners associated glossed words with sounds or images in dual modes at a time, they would learn the words better and retain them easier than when they used only one mode (Paivio, 1991). As such, Reiber (1994) states that "use of pictures and illustrations associated with unknown words are effective instructional devices that are superior to words alone for memory tasks and will help L2 learners remember the words sooner and retain them longer" (p. 141). The dual association of verbal and visual modes/annotations is also effective, because "when one memory trace is lost, the other remains and is accessible" (Lin, 2009, p. 24). Consequently, DCT is best applied to L2 vocabulary learning environments where texts, audios, and videos/animations are used. In other words, exposing learners to multiple modalities of presentation such as printed text, sound, picture, and video/animations produces a language-learning environment with a significant effect on vocabulary learning (Al-Seghayer, 2016, 2001).

The Cognitive Theory of Multimedia Learning (CTML)

The cognitive theory of multimedia learning (CTML) is an instructional theory about how people learn from multimedia presentations (Mayer, 2014, 2005, 2001, 1997). This theory draws on Wittrock's (1974) generative theory and Paivio's (1986) dual coding theory. However, it takes a step beyond these two theories and gives the learner the role of "knowledge constructor who actively selects and connects pieces of visual and verbal knowledge" (Mayer, 1997, p. 4). The theory centers on the idea that learners attempt to build meaningful connections between words and pictures and processes them actively in long-term memory (Mayer, 2005, 1997).

Mayer's (2014, 2005) cognitive theory of multimedia learning is based on three assumptions: dual channels, limited capacity, and active processing. Dual channel suggests that humans have two separate information processing channels (auditory/verbal and visual/pictorial). The auditory/verbal channel receives information (such as spoken or written words, narrations, & sounds) through the ear; and the visual/pictorial channel processes the information (such as pictures, graphs, videos/animation clips, & on-screen texts) through the eyes. The second assumption (i.e., limited capacity) signifies that there is a limit on the amount of information that learners can process in each channel at one time; therefore, the information is organized into two separate mental models. Words are stored in a verbal mental model and images are stored in a visual mental model, as Mayer (2014) explains, "when an illustration or animation is presented, the learner is able to hold only a few images in the visual channel of working memory at any one time, reflecting portions of the presented material" (p. 49). The same is also true when learners are presented with a narration in that they can only hold a few words in the verbal channel of working memory at any one time (Mayer, 2014). Finally, in order to build a connection between verbal and visual representations and integrate them into the learner's existing knowledge, some cognitive activities should be processed in long-term memory and bring them back into the short-term memory. This cognitive process accounts for the third assumption (i.e., active processing). The active cognitive processes include:

> (1) Selecting relevant words for processing in verbal working memory, (2) selecting relevant images for processing in visual working memory, (3) organizing selected words into a verbal model, (4) organizing selected images into a pictorial model, and (5) integrating the verbal and pictorial representations with each other and with relevant prior knowledge activated from long-term memory. (Mayer, 2014, p. 54)

Figure 2.2 illustrates how learners learn L2 vocabulary in a multimedia-based setting.

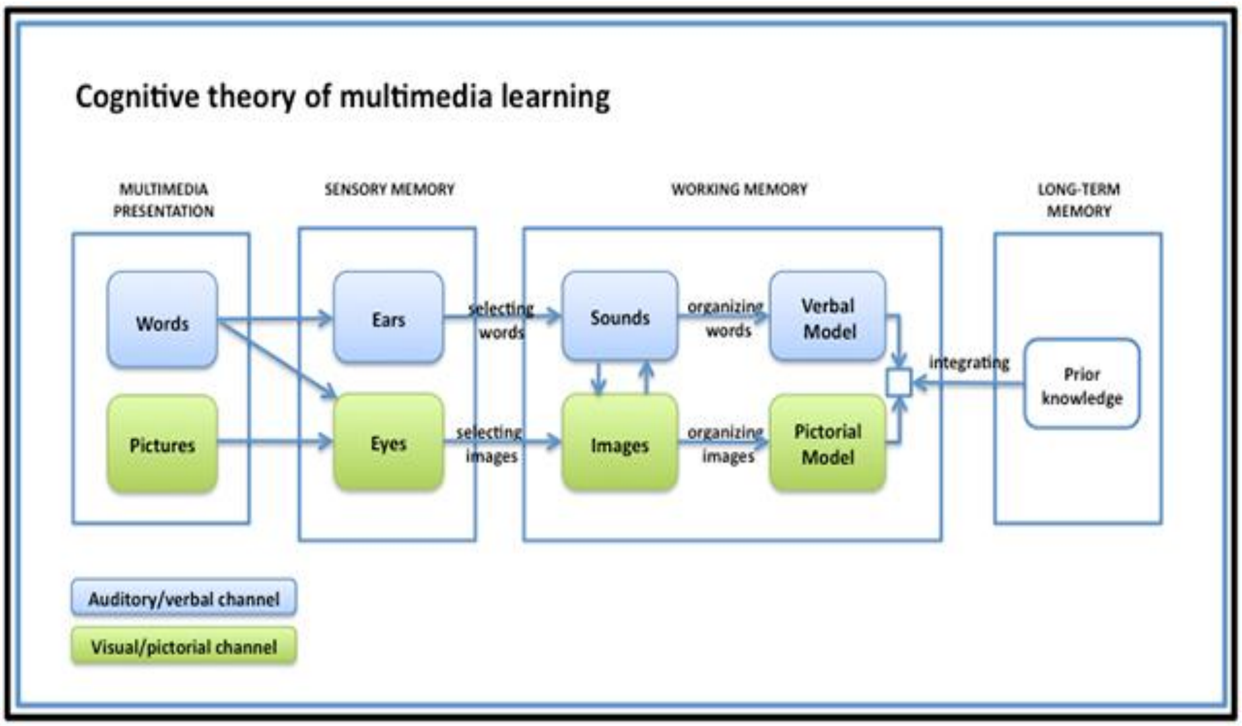


Figure 2. 2 Mayer's Cognitive Theory of Multimedia Learning (2014, p. 52)

According to the figure, words and pictures, as the two multimedia presentation modes, come to the sensory memory from the outside world through the ears and eyes. Words and pictures are kept in the visual sensory memory (**the green box**), and spoken words and sounds are held in the auditory sensory memory (**the blue box**) for a short time. In other words, words are stored in a verbal mental model and images are stored in a visual/pictorial mental model. Learners can select materials through attention to appropriate words and images. When relevant materials are selected, structural relations are built among the elements in working memory. The information is then transferred to the working memory where materials are temporarily held and manipulated. Working memory consists of two sides: the left side, which represents "the raw material" such as "visual images of pictures and sound images of words"; and the right side, which represents "the knowledge constructed in working memory" such as "pictorial and verbal models and the links between them" (Mayer, 2014, p. 53). The last box on the right side of Figure 2.2 shows long-term memory, which can hold large amounts of information over long periods of time. However, in order for the materials to stay in long-term memory, they should be actively moving back and forth from long-term memory to working memory (Mayer, 2014, 2005). In this way, knowledge in the long-term memory can be activated and brought into working memory if there exists a connection between new material and the learners' prior knowledge (Mayer, 2014).

In general, I drew on CTML in this study to provide support for the effectiveness of simultaneous multimedia glossing on L2 learners' vocabulary learning in respect to short and long-term word retention. I sought to justify that learners were more apt to recall vocabulary items when they had both verbal and visual formats available at the same time rather than one of these formats alone. The presence of these sources of information helped L2 learners to establish a direct mental connection between visual and verbal models in short-term memory, and paved the way for effective retrieval of words stored in the long-term memory (i.e., cognitive processes) (Mayer, 2014, 1997). Thus, having two separate but interrelated verbal and visual systems allowed the learners to benefit even more if they received the target words through the verbal tools of L2 text[8] and audio as well as the visual tools of text and video/animation in a multimedia-based learning environment. The two principles of multimedia learning, namely multimedia principle (Fletcher & Tobias, 2005; Mayer, 2005, 2001) and temporal contiguity principle[9] (Mayer & Fiorella, 2014) justify the rationale for the application of simultaneous verbal and visual glossing tools in this research. The next section briefly discusses the implications of the multimedia principle and the temporal contiguity principle in relation to the present study.

The multimedia principle. The multimedia principle suggests that learners can learn more effectively if they are presented with words and pictures rather than words alone (Mayer, 2014). However, the multimedia principle is not limited to words and pictures alone but refers to a broader term encompassing different forms of visual and verbal representations when presented together (Butcher, 2014, 2006). Visual components of the multimedia instruction include illustrations, pictures, graphs and charts, photographs, and videos/animations (Butcher, 2014). The verbal components are texts, spoken words/sounds, and narrations (Mayer, 2014). In this study, I employed the two glossing modes of L2 definition and audio glossing (text & spoken words) as well as L2 definition and video/animation glossing (text & picture) as the two verbal and visual components, along with L2 definition alone (text only) to present the target glossed words to L2 learners. Thus, according to the multimedia principle and its contribution to vocabulary learning, it was assumed that L2 learners could learn the target words better and more effectively when presented in dual modes rather than single mode, because the

[8] Text refers to L2 definitions of the target words.
[9] Review research rationale in introduction chapter for temporal contiguity principle.

use of both words and pictures allows the brain to process more information in working memory (Sweller, 2005), and can be recalled from long-term memory when required.

The temporal contiguity principle. This principle is designed to reduce extraneous overload of multimedia materials. It suggests that learners can learn more deeply from learning tools when the text, audio, pictures, and video/animation are presented simultaneously rather than successively or sequentially (Mayer & Fiorella, 2014). According to the cognitive theory of multimedia learning (Mayer, 2008), "learners must have corresponding words and images in working memory at the same time in order to make connections between them" (p. 764), meaning simultaneous presentation is expected to result in better learning than successive and separate presentation (Rusanganwa, 2015). The theoretical implication of the temporal contiguity principle for word learning in this study was that when L2 learners were presented with L2 definitions of the glossed words before or after they heard the audio recordings/spoken words, they had to hold all of the relevant target words in the working memory until the audio narrations/spoken words were presented. The same happened when the L2 definition of the target words was presented before or after the corresponding video/animation clips. Consequently, the task resulted in overloading learners' cognitive capacity. One possible way to eliminate the need to hold the information in working memory for a long time was to present the multimedia materials at the same time (Mayer & Fiorella, 2014). This way, the definitions and audio recording or the definitions and video/animation clips that were shown together, without keeping them separate, were more likely to contribute to L2 learners' word learning, reduce loads of material and recall than the presentation of L2 definition first and audio recording later (or L2 definition first and then video/animation) (Mayer & Sims, 1994).

Summary of Part One: Integrating the Two Theories

The theoretical constructs of dual-coding (DCT) and cognitive theory of multimedia learning (CTML) framed this investigation appropriately due to the overlap between the two, and their application to the domain of language learning, in general, and L2 vocabulary learning, in particular. The two theories unanimously imply that learning in a multimedia learning environment is best facilitated when the new information is presented through both verbal and visual representational modes rather than just a single mode (Mayer, 2014; Paivio, 1986). In other words, L2 learners' interactivity with multimedia input is enhanced through connecting both visual and verbal systems to written and pictorial cues in brain. Furthermore, utilizing multimedia annotations/glossing to instruct unknown words can help learners to experience word learning through the cognitive processes of selecting relevant words and images, organizing them into verbal and visual representations, and finally integrating the words into corresponding verbal and visual stimulus (Mayer, 2014). To minimize the chances of overloading a learners' cognitive system, both verbal and visual materials can be displayed simultaneously rather than successively (Mayer & Fiorella, 2014). Consequently, L2 learners in this study had the advantage of receiving input from a variety of resources that included both verbal (L2 definition & audio) and visual (L2 definition & video/animation) vocabulary information. The vocabulary information was thus presented to them at the same time in order to help them learn L2 words effectively and retain them longer. Learners' extraneous loads of materials were also reduced to the extent possible through developing rich instructional materials that imposed less heavy working memory load on the learners to process the information (Pass & Sweller, 2014).

Part Two: Literature Review

In this part, I present the literature relevant to this study in three sections: (A) vocabulary learning, (B) vocabulary glossing, and (C) gaps in the literature and conclusion. The first section consists of research on the following themes: (a) importance of vocabulary learning in language education; (b) two common types of vocabulary learning, namely incidental and intentional/explicit word learning; (c) vocabulary learning and memory retention; and (d) vocabulary learning strategies (VLS) and word retention. In the second section, I discuss (a) vocabulary glossing; (b) gloss definition; (c) gloss advantages; (d) gloss categories; and (e) gloss types/combinations. I then present the relevant current studies in the field and explain the gaps in the literature regarding this study with some concluding remarks in the last section. The chapter ends with a summary on both theoretical framework and literature review sections.

Section A. Vocabulary Learning

Importance of vocabulary learning in Language Education

Vocabulary knowledge is viewed as one of the main factors necessary for mastering another language (Schmitt, 2008). When a FL/L2 learner learns a new language, he/she needs to reach a certain vocabulary threshold to be able to develop linguistic abilities to use the language (Lomicka, 1998). The learner also needs to master the four language skills (listening, reading, writing, and speaking). Thus, vocabulary knowledge is fundamental to these skills, and a lack of vocabulary knowledge affects all four language skills (Gass, 1999; Zhang & Li, 2011). Paribakht and Wesche (1999) point out that the acquisition of vocabulary is a growing and recursive process that involves the integration of various kinds of knowledge along with gaining different levels of ability to make use of that knowledge in communication. In a similar vein, Nation (1993) explains that "vocabulary knowledge enables language use, language use enables the increase of vocabulary knowledge, and knowledge of the words enables the increase of vocabulary knowledge and language use and so on" (p. 6). Furthermore, knowing a word entails the ability to both recognize the word (receptive knowledge) and use it correctly (productive knowledge). Nation (2010) refers to this word knowledge as "the ability to recognize the spoken and written forms as well as the meaning of a word" (p. 47). With regard to the importance of vocabulary in language learning, McCarthy (1990) states that "it is the experience of most language teachers that the single biggest component of any language course is vocabulary" (p. 1). Without words to express a wide range of meaning, communication in L2 cannot happen in any meaningful way (McCarthy, 1990). However, vocabulary learning is often considered to be "the least well catered of all aspects of learning a language" (McCarthy, 1990, p. 1).

Vocabulary learning not only touches language mastery in a positive way, but it is also regarded as one "sub-goal of a range of goals" (Nation, 2007, p. 1) in the language classroom. Other goals include learning the pronunciation of words, the grammatical constructions of a language, the rules of conversational discourse, and the culture of the target language (Nation, 2007, 2001). FL/L2 learners also need to know a large number of lexical items in order to operate in another language (Schmitt, 2008; Nation, 2001).

However, one way to set up vocabulary learning goals for language learners is to work out how many useful words they need to know in order to utilize English (Schmitt, 2008; Nation, 2006). Schmitt (2008) refers to word coverage as “the percentage of lexical items in written or spoken discourse that a learner must know” (p. 330). Previously, Laufer (1989) believed that approximately 95% coverage was enough; whereas, Hu and Nation (2000) suggested that this figure is closer to 98-99% for written discourse. Yet, it is not clear if the same 98% figure is appropriate for dealing with spoken discourse (Schmitt, 2008). The word coverage of 98% means that one word in 50 is unknown, which still impedes comprehension (Carver, 1994) and is “probably a reasonable minimum coverage figure” (Schmitt, 2008, p. 330). Some studies have indicated that considerable comprehension in oral communication can occur with lower coverage rates of 95% (Bonk, 2000) or a coverage as low as 90% (Schmitt, 2008). Nation (2006) studied 200,000 words of the Wellington Corpus of Spoken English from interviews and friendly conversations between family members and friends to talk-back radios. With the use of “word lists based on (the mainly written) British National Corpus” (Schmitt, 2008, p. 330), Nation calculated that 6000-7000 word families are required in order to reach the 98% goal.

The scope of vocabulary learning, and “the fact that many learners fail to achieve even moderate vocabulary learning goals” (Schmitt, 2008, p. 333), indicate that although multiple exposures to FL/L2 words help establish a form-meaning connection and consolidate meaning (Nation, 2001; Schmitt, 1998), mere exposure to a second language is no longer assumed a “principled approach” to vocabulary learning (Schmitt, 2008, p. 333). Other vocabulary learning techniques such as learners’ engagements with words, time spent on learning a lexical item, and intentional/explicit vocabulary learning activities (Schmitt, 2008; Laufer, 2005) can help promote vocabulary learning. One way to achieve the form-meaning link is through intentional/explicit vocabulary learning (Schmitt, 2008), which is the next section's topic of discussion and the focus of this study.

Intentional/Explicit Vocabulary Learning

Intentional/explicit vocabulary learning is an approach to vocabulary instruction and learning (Hulstijn et al., 1996; Hunt & Beglar, 1998; Schmitt, 2008), where the lexical items of the target language are learned with an explicit focus on form and meaning (De Ridder, 2002; Khezrlou & Ellis, 2017; Schmitt, 2008) by means of synonyms, antonyms, word substitution, multiple choice items, scrambled words, and crossword puzzles (Akubulut, 2007). The intentional/explicit vocabulary learning, according to Schmitt (2008), always leads to "greater and faster gains, with a better chance of retention and of reaching productive levels of mastery" (p. 341). In intentional/explicit vocabulary instruction, new words that FL/L2 learners need to know are first diagnosed and then presented to the learners for the first time (Hunt & Burglar, 1998). Intentional/explicit vocabulary learning stands in opposition to incidental/implicit vocabulary learning, where FL/L2 learners are provided with adequate opportunities for extensive reading (Choi, 2016; Chun, Choi, & Kim, 2012; Hunt & Beglar, 1998; Nation & 1999; Waring & Takaki, 2003); thus, they are given the chance to guess the meaning of new words from the contextual clues (Ahmad, 2011; Paribakht & Wesche, 1999). In incidental/implicit vocabulary learning, the focus of the learning activity is not on learning words (Huckin & Coady, 1999), rather, it is "on understanding the message of the oral or written exchange” (Akbulut, 2007, p. 501). In other words, incidental vocabulary learning requires L2 learners to acquire unknown target words unintentionally as a result of engaging them in other learning activities, such as reading for pleasure or for gaining information (Choi, 2016; Choi, Kim, & Ryu, 2014; Huckin & Coady, 1999) or for language use, and focusing learners’ attention to the message to be conveyed rather than the linguistic structure (Wesche & Paribakht, 1999); whereas, intentional/explicit vocabulary learning is prompted when learners are pre-informed of a retention test after completing a reading task (Laufer & Hulstijn, 2001; Huckin & Coady, 1999), and the goal is to acquire the form and meaning of a word (De Ridder, 2003). The deliberate/explicit vocabulary learning, as well as incidental word acquisition, can each lead to the mastery of word knowledge (Barcroft 2009; File & Adams 2010; Hulstijn, 1992; Nation, 2013, 2001); however, the findings of studies show that deliberate learning results in more vocabulary learning, and “is an effective means of increasing vocabulary size” (Nation, 2011, p. 535) than incidental learning. Nation (2011) believes that explicit word learning is probably best “for learning the more salient aspects of word knowledge, particularly the form-meaning link” (p. 535). However, according to Schmitt (2002), mere deliberate/explicit learning may rarely "provide the knowledge of grammar, collocation, associations, reference, and constraints on use that may be best learned through meeting items in context" (p. 42); thus, contextual aspects of word knowledge such as collocations can be learned through incidental/implicit vocabulary learning (Nation, 2011).

In general, vocabulary learning results in developing techniques and approaches that assist learners in gaining a large number of words; and then assisting them to retain the words longer. Since the purpose of the present study is to first investigate the effectiveness of one of the vocabulary learning strategies, namely glossing, to enhance L2 learners' vocabulary learning, and second, to foster their short and long-term word recollection, intentional/explicit vocabulary learning was emphasized here as a broad vocabulary learning approach to consolidate form-meaning relationship, and to aid L2 learners to remember and use the unknown words over a longer period of time. Furthermore, research has shown that L2 vocabulary learning through extensive reading "often suffers from small gains, slow process, and requires repeated encounters" (Yoshii, 2014, p. 19); therefore, adding an intentional component such as glosses in a reading activity can encourage language learners to deliberately learn new words (Yoshii, 2014) and can promote greater efficiency to incidental vocabulary learning (Khezrlou & Ellis, 2017; Laufer, 2005; Schmitt, 2008; Wesche & Paribakht, 2000).

Focus of this study: Intentional/explicit vocabulary learning, in the present study, was defined as exposing L2 learners to the unknown words in the reading passages by providing them with L2 definitions using different glossing types. Besides, as L2 learners' proficiency level and learning situation are important in adopting any vocabulary learning/teaching approaches, highlighting intentional/explicit vocabulary instruction works best for intermediate L2 learners of this study who may have limited vocabulary repertoires.

Vocabulary Learning and Memory Retention

Merriam-Webster defines memory as the power or process of recalling what has been previously learned or experienced (www.merriam-webster.com/dictionary/memory). However, Preston (2007) states that memory is the mental capacity or brain's ability to recover, retrieve, and remember past events, impressions, and facts, and there are three main phases in the formation processing and retrieval of memory: (a) encoding or registration, which entails receiving, processing, and combining of received information; (b) storage of information, which encompasses creating a permanent record of the encoded information; and (c) retrieval, recall or recollection, which includes retaining the stored information in response to some cues for use in a process or activity (Preston, 2007). The next section first discusses different memory types and then relates memory retention to vocabulary learning.

Memory types. Zhang (2004) states that depending on the amount of time the memory lasts, memory is divided into three types: sensory memory, short-term memory, and long-term memory. Sensory memory is "the shortest-lived memory that lasts for milliseconds to a few seconds" (Zhang, 2004, p. 1). However, when the information lasts from several seconds to a few minutes, the memory is then called short-term memory (Zhang, 2004). Preston (2007) refers to short-term memory as primary/active/working memory. Short-term memory keeps the information which is already being processed, such as a new word encountered for the first time (Baddeley, 2002). Short term memory is fast, but has a very limited capacity to hold information in an active and readily available state for a very short time (Amiryousefi & Ketabi, 2011). In contrast, long-term memory is the memory which lasts anywhere from an hour to a lifetime (Zhang, 2004). Sweller and Chandler (1994) indicate that long-term memory has an unlimited storage capacity and can hold an indefinite amount of information provided that some changes happen in the long-term memory: The learning of any material, such as vocabulary, is the result of a change in a learner's long-term memory. Thus, appropriate alterations to long-term memory should be the primary aim of instruction in vocabulary learning environments (Pass & Sweller, 2014; Sweller, 2005); however, long-term memory processing is relatively slow.

Vocabulary and retention. Research has shown that, first, there is a close relationship between human memory and its ability to retain and recall information (Ellis, 1997; Stevick, 1996). The information, stored in the sensory memory, can be transferred from the short-term memory to the long-term memory if adequate attention is given, enough time is spent, and the information is rehearsed through a consolidation process (Mayer, 2014). In addition, relating new information to old ones signifies that a link is created between the two so that the new information can be incorporated into memory (Goodbridge, 2010). Second, memory has a crucial impact on eventual vocabulary learning and achievement (Amiryousefi & Ketabi, 2011). Learning a new word means establishing a relationship between form and meaning. This relationship may take the form of first language (L1) equivalents; L2 synonyms or in-text definitions; sounds, visual images such as static pictures or dynamic videos/animations; feelings and emotions; a certain situation or context; or a combination of these (Ahangari & Abdollahpour, 2010; Amiryousefi & Ketabi, 2011; Chen, 2016; Chun & Plass, 1996; Ko, 2005, 2012; Miyasako, 2002; Yoshii, 2006, 2014; Xu, 2010; Yeh & Wang, 2003, to name just a few).

Therefore, in order to transfer information accurately from working-memory to long-term memory, FL/L2 learners need to treat the information actively rather than passively, and interact with the information in meaningful ways (Schmitt, 2000). FL/L2 learners also need to look for both relationships and differences between the new information and other information that is already in long-term memory, and link them together (Amiryousefi & Ketabi, 2011; Mayer, 2014). One way to transfer the new lexical terms from the short-term memory to the long-term memory is to build a connection through finding some elements “in the mental lexicon” (Amiryousefi & Ketabi, 2011, p. 179), and attach the new lexical item to those elements (Schmitt, 2000). Information transfer in the present research context referred to the transfer of target words from L2 learners’ short-term memory to their long-term memory. Thus, learners required some vocabulary learning strategies to acquire new lexical information and transfer them to memory for consolidation purposes. The subsequent section provides an overview of vocabulary learning strategies in relation to word retention.

Vocabulary Learning Strategies (VLS) and Word Retention

Knowing that vocabulary learning is at the heart of any language learning and use (Laufer, 1997) signifies that it deserves comprehensive research. Moreover, retaining a large number of new words in memory has always been a challenging task for FL/L2 learners. Yet, no clear understanding has been accomplished on what vocabulary learning strategies can best help FL/L2 learners learn new words and remember them for later use. The reason might be that vocabulary learning has complicated and multifaceted strategies influenced by a wide variety of factors (Alessi & Dwyer, 2008; Gu, 2003; Groot, 2000; Nation, 2001). These factors include: (a) FL/L2 learners’ individual differences such as attitudes, beliefs, and motivation; (b) FL/L2 learners’ language learning experiences such as gender, field of study, and course type; and (c) FL/L2 learners’ learning outcomes like language proficiency, language achievement, and vocabulary knowledge (Boonkongsaen, 2012). Therefore, it is not surprising that FL/L2 teachers and learners are still uncertain of the best possible ways to follow vocabulary learning strategies in language classrooms (Schmitt, 2008). Moreover, retaining a large number of new words in memory has always been a challenging task for FL/L2 learners.

The word “strategy” is taken from the ancient Greek word “strategia,” which means steps or actions taken for the purpose of winning a war (Nemati, 2013, p. 9). Oxford (1990) defines language learning strategies as, "steps taken by students to enhance their own leaning” (p. 1). In other words, whatever learners apply to make learning easier, faster, and more enjoyable are regarded as language learning strategies (Oxford, 1990). Vocabulary learning strategies, as one subcategory of language learning strategies (Nation, 2001), are thus learning approaches which facilitate vocabulary learning (Amiryousefi & Ketabi, 2011); lead FL/L2 learners to take responsibility for their own learning (Ghorbani & Karimi, 2011); help them become active participants in the learning process (Nemati, 2013; Williams & Burden, 1997); and direct the learners both in discovering the meaning of a word, and consolidating it (Celce-Murcia, 2001). VLS are significantly stressed by some researchers and their usefulness has been widely examined (Amiryousefi & Ketabi, 2011; File & Adams, 2010; Min, 2008; Mizumoto & Kansai, 2009; Nation, 2001; Rott, Williams, & Cameron, 2002; Schmitt & Schmitt, 1995). With regard to vocabulary instruction, the findings of some studies indicate that individuals who use various vocabulary learning strategies could perform better in learning words (Lawson & Hogben, 1996; Rodiguez & Sadoski, 2000) and have longer word retention (Amiryousefi & Ketabi, 2011; Gu, 2003; Riazi & Alvari, 2004) compared to those who only memorize the words.

Schmitt (2008, 1997) divides the vocabulary strategies into five main types:

a. *Determination strategy* is used when learners analyze the word's meaning by looking at its structure, and parts of speech;

b. *Social strategy* is applied when L2 learners interact with others to find the meaning such as asking a teacher, a friend, or a native speaker for translation or meaning;

c. *Memory strategy* helps FL/L2 learners to memorize the information by relating it to their own experiences, images, meaningful context, videos/animations, grouping words with the same meaning together, and so forth;

d. *Cognitive strategy* includes repetition of the words either orally or by writing, taking notes in class, or labeling things; and finally

e. *Metacognitive strategy* entails self-testing, skipping unknown words, and enhancing vocabulary knowledge by using L2 media.

Nevertheless, among the above-mentioned strategies, I adopted memory strategy as one vocabulary learning strategy to help L2 learners to learn the target words effectively and remember them longer. Schmitt (1997) stressed that "most memory strategies (traditionally known as mnemonics) involved relating the word to be retained with some previously learned knowledge, using some form of imagery, or grouping" (p. 211). Examples of memory strategies that can promote, strengthen, and speed up vocabulary learning and word retention are the key word method (Tavakoli & Gerami, 2013), using a dictionary (Knight, 1994; Nemati, 2009), and glossing (Bowels, 2004; Gettys, Imhof, & Kautz, 2001; Cheng & Good, 2009; Hulstijn, 1992; Jacobs et al., 1994; Ko, 2005; Nagata, 1999; Watanabe, 1997).

Given the important role of vocabulary learning in FL/L2 education, both in oral and written language comprehension (NICHD, 2000), it is vital to define vocabulary learning. Gu (2003) views the task of vocabulary learning in two ways: (a) through the distinction between “knowing a word and using a word” (p. 3); and (b) through seeing the distinction “as a process of related sub-tasks” (p. 3). As for the first view, because the purpose of vocabulary learning is to assist FL/L2 learners in both remembering words and using them in a wide range of language contexts when the need arises (McCarthy, 1984), language educators and researchers should develop vocabulary learning strategies that help the learners to use and know the words. As for the second view, FL/L2 learners might use different tasks for vocabulary learning that determines to a large extent how well a new word is learned (Gu, 2003). These tasks include guessing word meaning and usage from available clues, namely meaning inferences, and extensive reading (Schmitt, 2008); looking words up in a dictionary; note taking along the margins, between the lines, or on separate vocabulary notebooks; repeating the words several times; and using the words actively in communicative occasions and glossing. The present study focuses on glossing as one practical and instructional means of the memory-aiding strategy. Glossing is useful because it provides accurate meanings for words that might not be guessed correctly (Nation, 2001). It has minimal interruption to reading when compared to the dictionary use (Lenders, 2008; Nation, 2001) and draws attention to words that should aid the acquisition process (Nation, 2001; Schmitt, 2008; Yanguas, 2009); thus, it requires further attention. To this aim, the focus of the second section of this chapter is directed towards vocabulary glossing; its advantages and possible types/combinations.

The section further discusses how glossing can assist FL/L2 learners in learning new words effectively and retaining them longer.

Section B: Vocabulary Glossing

Research has shown that FL/L2 learners are faced with the challenging task of remembering many new words (Scafidi-Iannone, 2012; Stewart & Cross, 1991), but still how learners can deal with this task is under investigation. Nation (2001) argues that "determining what strategies, within a classroom setting, help in the retention of vocabulary is important to the task of building vocabulary" (p. 1). Yet, there are some strategies which can promote L2 vocabulary learning, and increase word knowledge such as dictionary use (Knight, 1994), and glossing (Jacobs et al., 1994). Several studies have examined the effectiveness of glosses for L2 vocabulary learning, and the findings revealed that glossing enhances L2 vocabulary learning (Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b; Hong, 2010; Hulstijn et al., 1996; Hulstijn, 1992; Jacobs et al., 1994; Knight, 1994; Kost et al., 1999; Samian et al., 2016; Watanabe, 1997). Gettys and colleagues (2001) found that glosses "provide fast and easy access to the meanings of unknown words" (p. 93) and help learners comprehend a text better. However, studies show inconsistent results on exactly what gloss combination(s) best promote L2 vocabulary learning and enhances word retention (Akbulut, 2007; Al-Seghayer, 2001; Ariew & Erçetin, 2004; Chun & Plass, 1996 a, 1996b; Lin & Tseng, 2012; Plass et al., 1998; Sakar & Erçetin, 2005). Therefore, further research is required to examine how glossing can develop L2 vocabulary learning of FL/L2 learners and assist them in remembering the words effectively. In the following sections glossing is first defined and the advantages are elaborated. Second, while giving a short overview on gloss categories and types/combinations in the field of L2 vocabulary learning, for the purpose of consistency and relevance to this research, I expand the literature on multimedia glossing to demonstrate that multimedia glossing provides FL/L2 learners with different modalities (textual, visual, and aural) and modes of learning (texts, pictures, and video/animations). Further, I explore the findings of some studies in relation to the applicability of multimedia glossing in L2 vocabulary learning and word retention.

Gloss Definition

The concept of glossing dates back to the middle ages when learners had problems with understanding unfamiliar texts, such as Latin (Shahrokni, 2009). Glossing was traditionally referred to as a short definition or note to facilitate reading comprehension for L2 learners (Lomicka, 1998). However, glosses have been defined differently depending on their various functions, which help learners "to decode the text by providing additional knowledge in specific content, skills, strategies and definitions of difficult words" (Ko, 2005, p. 125). As an example, Nation (1983) defined glosses as short definitions; Pak (1986) mentioned them as explanations of the meanings of words; and Lomicka (1998) referred to glossing as "typically located in the side or bottom margins" (p. 41) and defined it as "the most often supplied form for "unfamiliar" words, which may help to limit continual dictionary consultation that may hinder and interrupt the L2 reading comprehension process" (p. 41). AbuSeileek (2008) also refers to glossing as "adding comments or notes about difficult words, phrases, or ideas in order to provide their definitions or meanings in a particular context" (p. 260). Lomicka's (1998) interpretation of glossing implies that glosses are effective no matter where they are located, and how they are used, and they may bear different types of information such as cultural, historical, geographical references, and guiding questions (AbuSeileek, 2008; Lomicka, 1998). Moreover, Stewart and Cross (1991) define glosses as a “bridge between learners’ prior knowledge and new information” (p. 5), whereas Roby (1999) rephrases the bridging definition of glossing as “metacognitive gloss” (p. 95) to improve learners’ reading comprehension. Finally, Roby (1999) refers to glossing as various kinds of “attempts to supply what is perceived to be deficient in a reader’s procedural or declarative knowledge” (p. 96). Procedural (skill) knowledge refers to the demonstration of knowledge through the procedure of doing something, while declarative (factual) knowledge refers to the factual information that a person knows. Roby (1999) emphasizes that glosses are much more than translations or explanations of difficult words.

In sum, in the case of FL/L2 vocabulary learning environments, glossing generally refers to providing additional information on important, difficult, or technical words via definitions, examples, translations, and/or synonyms of unknown words either in FL/L2 learners’ first language (L1) or L2 (Nation, 2006; 2001), or at the end of the text or near the unknown words (Türk & Erçetin, 2014). In the light of research on glossing, it can be concluded that glossing is advantageous over traditional techniques of vocabulary learning, such as dictionary use, word-lists, and teacher materials, guessing, and inferring meaning from context, to name a few, for L2 vocabulary development and long-term word retention (Al-Seghayer, 2001; Cheng & Good, 2009; Farvardin & Biria, 2012; Hong, 2010; Hulstijn et al., 1996; Jacobs et al., 1994; Ko, 2012; Türk & Erçetin, 2014; Watanabe, 1997; Zandieh & Jafarigouhar, 2012). The next section further elaborates on the benefits of using glosses for word learning and recollection.

Gloss Advantages

Widdowson (1978) argues that glossing is important in FL/L2 reading comprehension, because it is an effective way for learners to understand what they read. In a similar vein, several studies have found that glossing not only helps FL/L2 learners to learn new words effectively (Al-Seghayer, 2001; Chun & Plass, 1996), but they also support the claim that it can enhance long-term word retention (Salem & Aust, 2007; Salem, 2006; Zandieh & Jafarigohar, 2012). Some of the advantages of glossing in FL/L2 vocabulary context are summarized below:

1. Glossing has been shown to help L2 learners understand new words more accurately by preventing incorrect guessing (Hulstijn, 1992; Ko, 2005; Nation, 2013, 2001). Hong (2010) indicates that "guessing the meaning of new words from context can be difficult and risky" (p. 60) if readers lack adequate knowledge about the language or the reading strategies. As an example, FL/L2 learners can best guess the meaning of unknown words only when they understand the context and know all the surrounding words in the text. Thus, glossing prevents learners from making wrong inferences (Nation, 2013, 2001).

2. Glossing can make learners less dependent upon their teachers, allowing for autonomy (Nation, 2002; Rott, Williams, & Cameron, 2002) or "independent readers" (Stewart & Cross, 1991, p. 11). Since not all FL/L2 learners have problems with the same words, they can just search for the words they do not know (Jacobs, 1994; Ko, 2005; Nation, 1990). Also, according to Stewart and

Cross (1991), “with glossed text, three voices become involved in the reading: the inner voice of the reader, the voice of the author, and the voice of the teacher manifested in the gloss” (p. 5). Moreover, they maintain that “the purpose of glossing is to produce independent readers” (Stewart & Cross, 1991, p. 11).

3. Glossing helps FL/L2 readers build a bridge between the prior knowledge and the new information in the text (Stewart & Cross, 1993, 1991). In this way, FL/L2 learners may develop and enhance their comprehension (Jacobs et al., 1994) and retention of the text content if there is an interaction among the gloss, the reader, and the text (Ko, 2005). This interaction helps FL/L2 learners remember their background knowledge and connect it to the text (Stewart & Cross, 1993). Because of the proximity of gloss to the text, learners are provided with minimal interruptions during the reading flow (Nation, 2013, 2006, 2001) as learners are provided with immediate access to the word definitions compared to looking up words in a traditional dictionary setting (Chen, 2016; Chen & Yen, 2013; Lenders, 2008; Nation, 2001); hence retention of the content of the text is improved (Ko, 2005; Stewart & Cross, 1993, 1991). Consequently, if properly used, glosses draw learners’ attention to words (Boers et al., 2017); thus, may encourage the learning of new vocabulary (Rott, 2007; Nation, 2001), saves learners’ time and effort (Nation, 2001), and caters for learners’ preferences (Jacobs et al., 1994).

4. Glossing can also provide exposure to unfamiliar and new vocabulary, and hence increase word retention (Al-Seghayer, 2001; Chun & Plass, 1996b; Hong, 2010; Ko, 2005; Kost et al., 1999; Lomicka, 1998; Nation, 2011, 2013; Plass & Jones, 2005; Yoshii, 2006). These multiple exposures include encountering unknown words for the first time in the text (first exposure), learning the words through different gloss modes, such as textual, aural, and visual (second exposure), and relating the meaning to the words in the text to make sure if it fits the content (third exposure) (Nation, 2013; Watanabe, 1997). Therefore, the back and forth movement between the text and the target words facilitates word learning and retention.

5. Glossing can activate learners' motivation to notice and attend to the target words based on the notion of consciousness-raising and input enhancement. Learners will be able to do lexical processing to retain the words by frequently referring to the glossed words (Nagata, 1999). Glossing also attracts learners' attention to reading the text (Nation, 2013, 2001) and creates high level of motivation in learning the unknown words (Zoi, Bellou, & Mikropoulos, 2011).

6. Since glossing provides definitions for low frequency words, L2 learners do not need to continually look them up (Nation, 2001, 1990). Glossing can also serve as a technique to substitute for a dictionary (Yanguas, 2009) and can provide greater use of authentic and un-simplified texts (Nation, 2013; Jacobs et al., 1994).

In conclusion, glossing not only has the purpose of aiding reading comprehension in a certain task, but leads FL/L2 learners to know about the benefits of looking up unfamiliar words when the purpose is to achieve full understanding of a text. Glossing can be a helpful and practical learning device to enhance lexical retention, and a vocabulary learning strategy that saves students' time and effort in reading L2 texts.

Gloss Categories

Researchers have referred to different taxonomies and classifications for glossing (Lomicka, 1998; Roby, 1999; Stewart & Cross, 1991; Widdowson, 1978). Stewart and Cross (1991) categorize glossing as elaboration, bridging, key point, and vocabulary glosses. They state that elaboration glosses "clarify the text with statements, questions, or both", whereas bridging glosses "combine a question and act as a bridge between prior learning and new information" (Stewart & Cross, 1991, p. 5). The key point gloss emphasizes important statements in a text, and the vocabulary glosses simply provide definitions of the unknown words. In a different classification, Roby (1999) recasts Stewart and Cross' (1991) category and presents a detailed taxonomy for glossing, which includes all possible gloss divisions. In his category, glosses are grouped according to authorship, presentation, focus, language, function, and form, as shown below.

I. *Gloss authorship*
   A. Learners
   B. Professionals
      1. Instructors
      2. Materials developers

II. *Gloss presentation*
   A. Priming
   B. Prompting

III. *Gloss focus*
   A. Textual
   B. Extra-textual

IV. *Gloss language*
   A. L1
   B. L2
   C. L3

V. *Gloss function*
   A. Procedural
      1. Metacognitive
      2. Highlighting
      3. Clarifying

VI. *Gloss form*
   A. Verbal

B. Declarative	B. Visual
1. Encyclopedic	1. Image
2. Linguistic	2. Icon
a. Lexical	3. Video
i. Signification	a. with sound
ii. Value	b. without sound
b. Syntactical	
	C. Audio (only)

Source: Gloss taxonomy, Roby (1999, p. 96)

The present study followed Roby's (1999) taxonomy of gloss form, language, and presentation. According to this category for the gloss form, the three modes of verbal (L2 definition), visual (video/animation), and audio glossing (sound recording) were used to present new words to L2 learners; English, as a second/foreign language of the learners was consulted for the target word definitions; and the priming glossary presentation was valued over prompting glossary mode. Widdowson (1978) has proposed that "glosses written by teachers or materials developers are termed as priming glossaries" (p. 82), and they provide the definitions of unfamiliar vocabulary items; whereas, "glosses that the reader consults in the course of reading" (p. 86) are labelled as prompting glossaries. Prompting glosses give the meaning of a vocabulary item in a particular context. Yet, Roby's (1999) gloss classification is comprehensive as it can represent various taxonomies and characterizations for vocabulary glossary instruction, from gloss presentation and function to gloss focus, language and form. This classification is also aligned with the objectives of the present study.

Generally, glossing can provide information on new words through pictures, adequate context, and videos (Nation, 2001; Stewart & Cross, 1993, 1991); sounds and icons/images (Lomicka, 1998); and interactive multimedia elements (Rouhi & Mohebbi, 2013; Al-Seghayer, 2001). The next section depicts different gloss types/combinations as presented in literature.

Gloss Types/Combinations

The application of different kind of glosses is one of the recent techniques in L2 vocabulary learning. There are various manifestations of information[10] presented in the gloss. Nation (2013) classifies glosses into hard-copy and electronic glosses. However, the type in which new words are presented varies from single-mode glosses such as textual definition alone, pictorial definition alone, audio alone (AbuSeileek, 2008; Boers et al., 2017; Hulstijn & Laufer, 2001; Marzban, 2011; Nagata, 1999; Rassaei, 2017; Salem & Aust, 2007; Tabatabaei & Shams, 2011; Yeh & Wang, 2003; Yoshii & Flaitz, 2002; Yusuf et al., 2014) to dual-mode/bimodal glosses such as textual and pictorial definitions, textual and video/animations (Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b; Kost et al., 1999; Lomicka, 1998; Yoshii, 2006); multimodal glosses like textual, pictorial, and aural definitions (Salem & Aust, 2007)[11], and first language (L1) translations versus L2 glosses (Choi, 2016; Lomicka, 1998; Jacobs et al., 1994; Jung, 2016; Ko, 2005, 2012; Xu, 2010). Furthermore, researchers have considered the impact of other gloss types that include marginal glosses (AbuSeileek, 2008; Chen, 2016; Hulstijn et al., 1996; Jacobs et al., 1994; Sadeghi & Ahmadi, 2012; Samian et al., 2016; Stewart & Cross, 1991); multimedia glosses (Ahangari & Abdollahpour, 2010; Al-Seghayer, 2001; Plass et al., 1998; Türk & Erçetin, 2014; Yanguas, 2009); and multiple-

---

[10] Here, information refers to the definition of the target/unknown words.

[11] All the studies referenced here have used one, two or several gloss modes, depending on their study design.

choice glosses (Rott, 2005; Nagata, 1999; Watanabe, 1997) on L2 vocabulary learning. In what follows, I have attempted to present the various gloss types and combinations and their contributions to L2 vocabulary glossing as introduced in the literature from the dated to the most recent ones chronologically.

In an early study, Hüllen (1989) remarked that glosses were, at first, divided into three major types: synonyms, encyclopedic comments, and grammatical notes, and they were utilized for the purpose of understanding Latin; whereas Landers' (2008) recent gloss types fall into the three forms of: (a) dictionary-type; (b) ready-made type; and (c) special electronic and non-electronic type. A brief description of each follows.

a. Dictionary-type gloss contains information about the meaning(s) of a word, including definition, synonym, antonym, L1 equivalent, phonetic script, or example sentences (Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b). Figure 2.3 illustrates a glossed-word accompanied by a color picture in a multimedia mode. The word is defined according to dictionary definition.

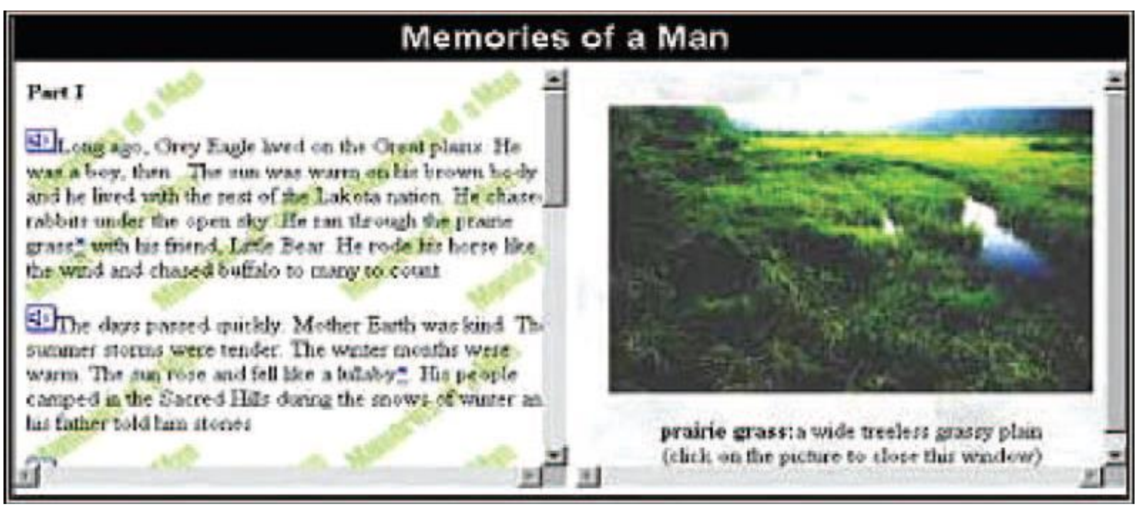


Figure 2. 3 Dictionary-type Gloss in Multimedia Environment (Al-Seghayer, 2001, p. 228)

b. Ready-made gloss type is specially made for the particular needs of learners in a course. The glosses provide information concerning the meaning of the glossed word in the given context. Ready-made glosses are used in studies which allow for teaching specialized/technical vocabulary in authentic contexts, thus they are widely used (Al-Ghafli, 2011; De Ridder, 2002; Gettys et al., 2000; Lenders, 2008). This gloss type may contain a spoken or written L2 definition, an L1 translation, or a still or moving image, and can be used exclusively or in any combination (See Figure 2.4).

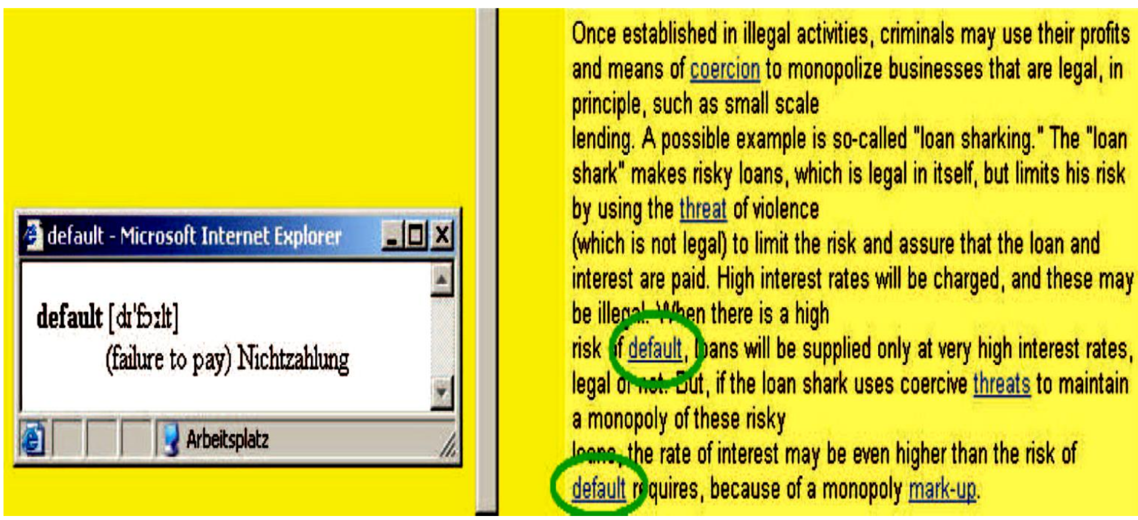


Figure 2. 4 Ready-made Gloss (Lenders, 2008, p. 458)

c. Special gloss types provide the definition of words in electronic or non-electronic format. What makes this gloss type different is that it includes "a task for the learner" (Lenders, 2008, p. 459) or provides more than one possible translation of the gloss word in multiple-choice format (Nagata, 1999; Watanabe, 1997). According to Lenders (2008), "if learners have to decide which of the given words provides the correct translation it leads to increased cognitive processing of the glossed word" (p. 459). As a drawback to this gloss type, he further explains that learners may learn the wrong meaning of the new words if they do not receive sufficient feedback to understand the correct meaning of the glossed words (Lenders, 2008).

However, within a broader perspective, as the relevant literature shows, glosses can also fall into single-mode, dual-mode/bimodal, and multimodal types. Each gloss mode is briefly explained in the next section.

Single-mode glossing. Single-mode gloss type provides only brief definitions or translations of unknown words in the margin or back of the text (Azari, 2012; Gettys et al., 2001; Watanabe, 1997; Yusuf et al., 2014). The single modes vary from text-only type to audio-only, picture-only, and or video/animation-only type glosses. Figure 2.5 illustrates a single-mode gloss type (text-only gloss), in that the learner is required to read the text and learn the highlighted word while referring to its dictionary-based definition or synonym only. Figure 2.6 also presents an example of picture-only glossing accompanied with the target word.

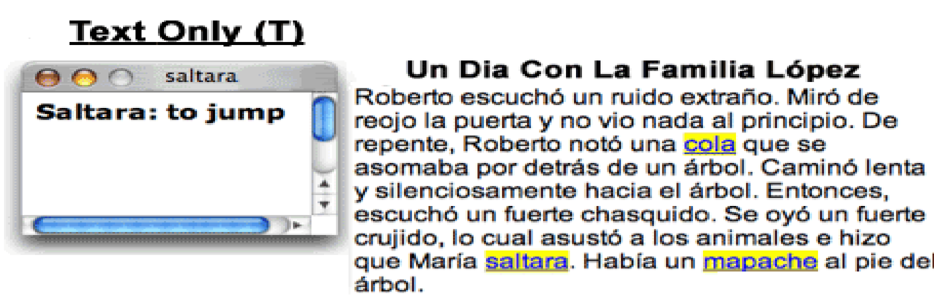


Figure 2. 5 Single-mode Gloss type (Text-only) (Salem, 2006, p. 54)

Figure 2. 6 Single-mode Gloss type (Picture-only-reunion) (Yeh & Wang, 2003, p. 135)

Dual-mode/bimodal glossing. Dual mode/bimodal gloss presents information through the combination of text and pictures (Chun & Plass, 1996a, 1996b; Yoshii & Flaitz, 2002), text and audios/sounds (Salem, 2006; Yeh & Wang, 2003), or text and videos/animations or dynamic pictures (Ahangari & Abdollahpour, 2010; Al-Seghayer, 2001; Lin & Tseng, 2012). Figure 2.7 shows dual-mode glosses of text and audio. Figure 2.8 displays bimodal glossing of text and picture.

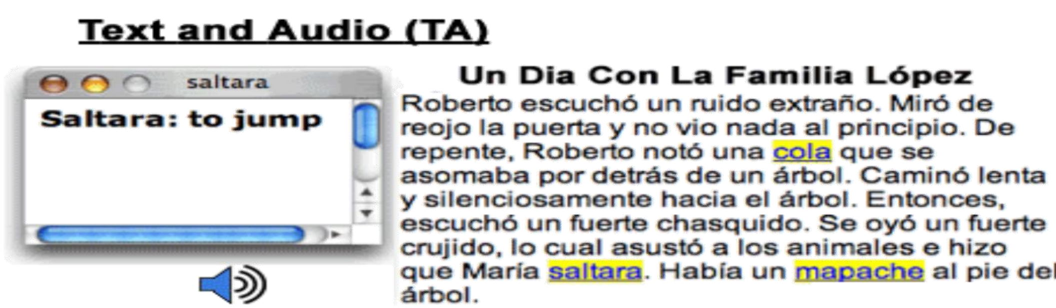


Figure 2. 7 Dual-mode/Bimodal Gloss type (Text & Audio) (Salem, 2006, p. 55)

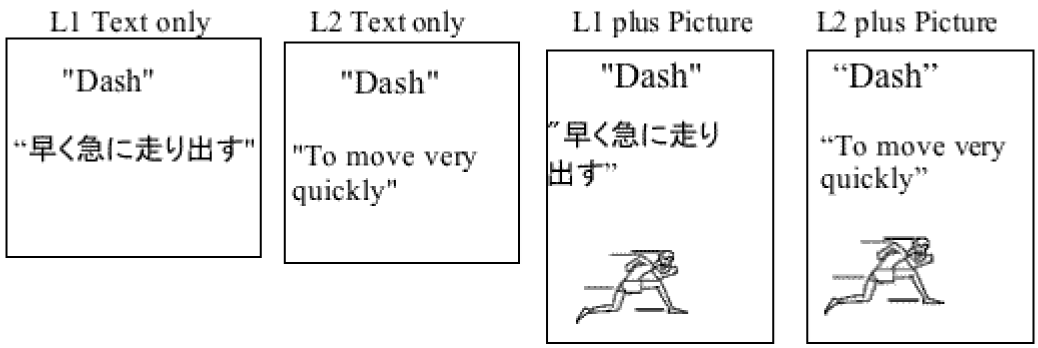


Figure 2. 8 Dual-mode/Bimodal Gloss type (Text & Picture) (Yoshii, 2006, p. 90)

Multimodal glossing. This gloss type refers to short definitions or explanations with graphics, audios, and videos/animations (dynamic pictures) in computerized or non-computerized texts (Salem & Aust, 2007; Bowles, 2004). Within the broad category of multimodal glossing, multimedia glossing emerges (Al-Seghayer, 2001; Erçetin, 2003; Lomicka, 1998; Salem, 2006; Türk & Erçetin, 2014). Multimedia gloss type presents explanations/information of target words via multiple types of resources such as texts, graphics, still pictures, sounds, and dynamic videos/animations in a multimedia-based learning environment such as the language lab or language classroom where a computer, screen, a projector, internet, and/or other technological tools are available. Multimedia glossing can also take the form of interactive multiple-choice glossing (Al-Seghayer, 2001; Laufer & Hulstijn, 2001; Türk & Erçetin, 2014), which displays multiple definitions for a new and unfamiliar vocabulary items in the text successively or simultaneously (Türk & Erçetin, 2014) (See Figures 2.9, 2.10 & 2.11).

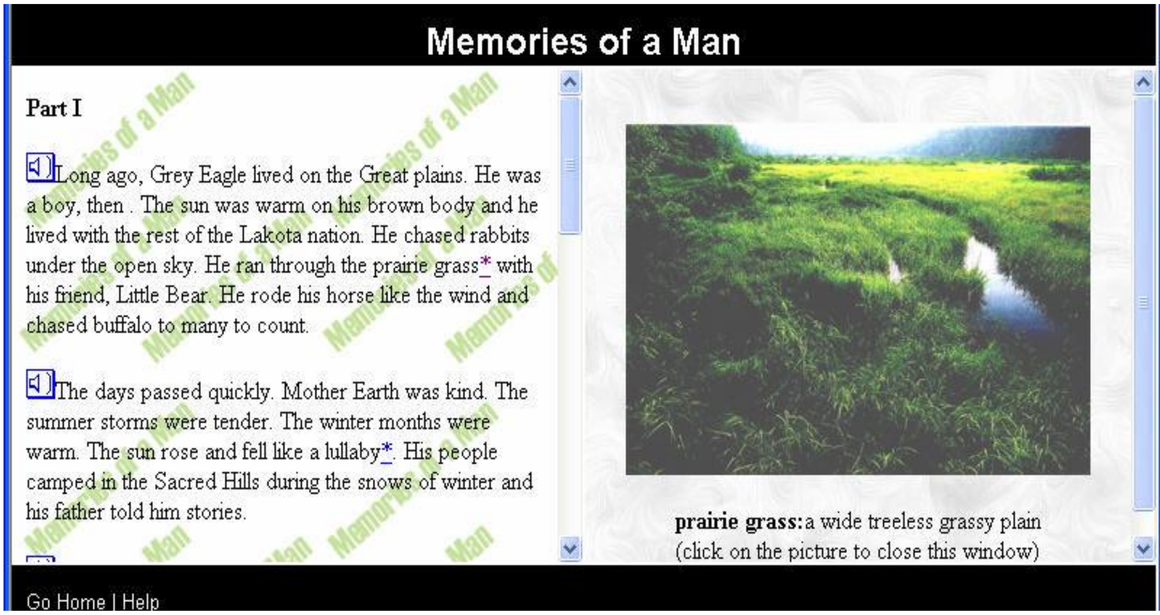


Figure 2. 9 Multimodal Gloss type (Audio, Text & Picture) (Al-Seghayer, 2003, p. 12)

Figure 2. 10 Multimedia Gloss type (Text, Audio, Picture & Animation) (Al-Ghafli, 2001, p. 2013)

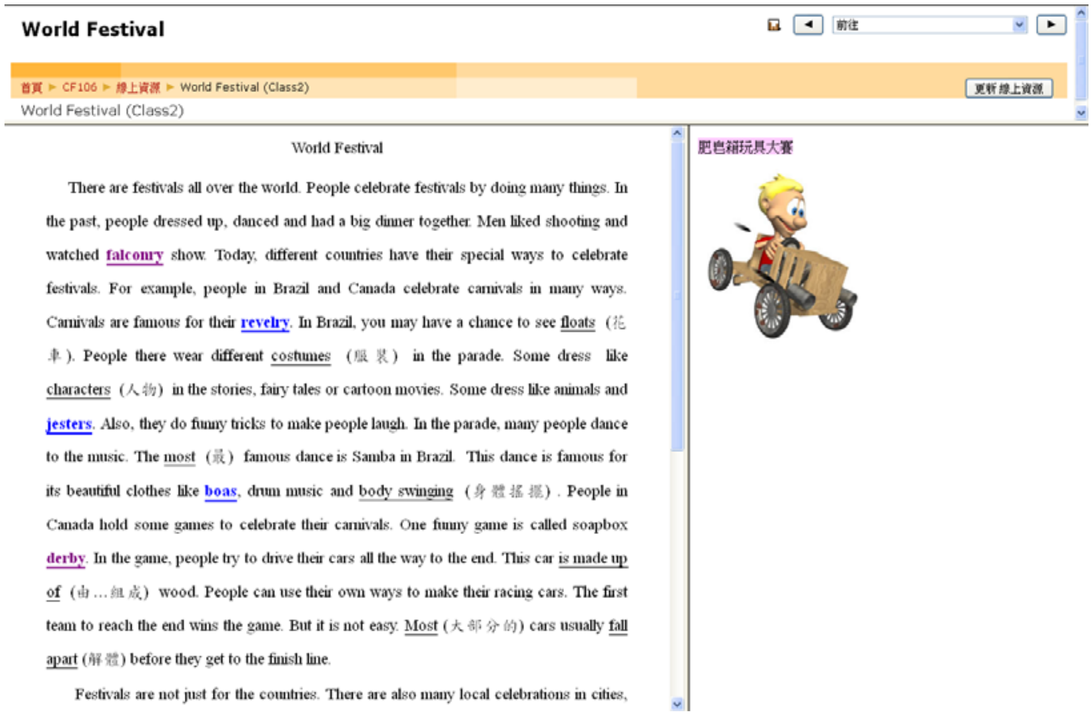


Figure 2. 11 Multimedia Gloss type (L1 definition alone, L1 definition & Picture, L1 definition & video/Animation) (Lin & Tseng, 2012, p. 349)

The focus of the present study was to investigate the effectiveness of multimedia glossing on L2 vocabulary learning in terms of short and long-term word retention. The research aimed to examine if multimedia gloss presentation would effectively enhance L2 learners’ vocabulary learning and word recollection in short and long-term. Accordingly, multimedia glossing was defined as textual definition of the target word alone (i.e., L2 definition), the combination of textual definition of the target word in L2 with its audio pronunciation, and the combination of textual definition of the target word in L2 with its relevant video/animation clip. In the subsequent section, multimedia and its advantages in vocabulary learning are first discussed, and the literature on the multimedia vocabulary glossing is then reviewed in order to explore the findings of some recent studies on this domain.

Vocabulary learning with multimedia. As stated earlier in the introduction chapter, Mayer (2014) defines multimedia as the presentation of both words and pictures to language learners. In his definition, words signify any material presented in verbal form such as spoken texts or printed texts; and pictures represent pictorial materials such as static graphics, illustrations, graphs, diagrams, maps, or dynamic graphics including animation or video. Thus, multimedia learning is defined as "building mental representations from words and pictures" (Mayer, 2014, p. 2). In other words, multimedia learning is a research-based, well-supported theory of how people learn from words, sounds, and images (Mayer & Moreno, 2003). Brinton, as cited in Sydorenko (2010), maintains that multimedia use is supported for the following reasons: (a) "it allows for the provision of authentic input and thus exposure to target culture, (b) it motivates learners, and (c) it takes account for students' different learning styles" (p. 50). The application of multimedia in language classes is not only effective for language learners to save time and effort in reading L2 passages and enhance vocabulary retention (Gettys et al., 2001), but also helps language teachers to facilitate language instruction, in general, and vocabulary instruction, in particular. As such, Devi (2010) states that "because of its capability of integrating the four skills of listening, speaking, reading and writing, multimedia is of considerable interest to the language teacher" (pp. 72-73).

A considerable amount of research has been conducted on the use of multimedia for vocabulary learning (AbuSeileek, 2011; Ahangari & Abdollahpour, 2010; Akbulut, 2007; Al-Seghayer, 2001; Chen & Yen, 2013; Mohsen, 2011; Mohsen & Balakumar, 2011; Plass & Jones, 2005; Plass et al., 1998; Rusanganwa, 2015; Salem, 2006; Türk & Erçetin, 2014; Yanguas; 2009; Yoshii, 2006). Some studies have provided evidence for the positive effect of multimedia use in L2 vocabulary learning. However, other studies lend support to the insignificant impact of multimedia glosses on vocabulary learning in the long-term (Jacobs et al., 1994; Black, Wright, Black, & Norman, 1992). As an example, Jacobs et al., (1994) found that glossing was positively more effective than non-glossing for learners' vocabulary acquisition in short-term; however, the impact did not last for learners' long-term purposes.

Multimedia/electronic Glossing: A survey in current literature. Multimedia/electronic glossing emerged as a result of the advancements in computer technology and media (Salem & Aust, 2007), and the integration of glosses into computer software programs (Abu Seileek, 2011; Akbulut, 2007; Al-Seghayer, 2001; Ariew & Ercetin, 2004; Chen, 2006; Chun & Plass, 1996; Kost, et al., 1999; Mohsen, 2011; Salem, 2006; Plass et al., 2003; Rusanganwa, 2015; Yanguas, 2009; Yeh & Wang, 2003) to improve L2 vocabulary learning. Multimedia/electronic glosses take the form of various vocabulary annotations in multimedia environments such as texts, audios, pictures, and videos/animations. Najjar (1996) remarks that the use of multimedia-based instruction in classroom education takes less time for the learners to learn the material and leads them to better organization and structure than traditional classroom lectures. In a similar vein, Mayer and Moreno (2002) maintain that computer-based multimedia learning environments provide an influential situation to improve learner's understanding of the subject matter. Given the positive role of multimedia instruction in language classes, there is a tendency to base glossing on computer to facilitate L2 vocabulary learning. Additionally, using multimedia glossing is an interactive (Ariew & Ercetin, 2004) and learner-oriented technique that assists FL/L2 learners in developing vocabulary repertoire and facilitating reading comprehension. Generally, there are several advantages to multimedia glossing: (a) enhancing reading comprehension; (b) increasing vocabulary learning; (c) providing FL/L2 learners with adequate use of authentic texts; (d) attending to learners' preferred mode of vocabulary learning; (e) improving vocabulary retention; and (f) saving learners' time and effort in reading L2 texts (Al-Seghayer, 2001; Ariew & Ercetin, 2004; Gettys et al., 2001; Jacobs et al., 1994; Khezrlou & Ellis, 2017; Mohsen & Balakumar, 2011). Furthermore, in favor of computer-based multimedia glossing, the results of the three meta-analyses (Abraham, 2008; Yun, 2011; Vahedi & Ghonsooly, 2016) found that there is a positive effect of computer-based glossing for word learning and retention. Abraham's (2008) study revealed a large effect size of computer-mediated glosses on incidental vocabulary acquisition. Thus, multimedia/electronic glossing deserves special attention in L2 vocabulary learning. In the following section, the effectiveness of multimedia glossing on L2 vocabulary learning and retention is further explored, and the controversies on various multimedia/electronic gloss combinations are discussed. It is noteworthy to mention that several studies have been conducted to date using different gloss combinations; however, for the purpose of relevance and consistency, only those studies which used glossing in a multimedia-based learning environment are discussed below. These studies are among the most cited research on multimedia glossing and L2 vocabulary learning.

An example of notable research in the category of multimedia glossing is Chun and Plass (1996a, 1996b) who conducted a series of studies that investigated the effectiveness of different gloss modes of textual/verbal and visual on vocabulary learning of a group of L2 learners enrolled in a German course. The researchers designed a program, called CyberBuch, which contained textual annotations in German reading texts (in English) and visual annotations (video and pictures). The participants were randomly assigned to three groups: text only, text and still-picture, and text and video. Results of this study revealed that firstly, participants preferred visual modes (video & still-pictures), and secondly, they acquired more words with annotations of text and picture than the other two annotations with text only or text and video annotations. In other words, words glossed with both text and picture helped the participants recall more vocabulary than words glossed with text and video. Furthermore, according to the findings of production and recognition tests, the annotated modes of text and picture as well as text and video were better than text annotations alone. The study also supported dual-coding theory, the rationale of which is that “learning of a vocabulary item is best when both visual and verbal information are present” (Chun & Plass, 1996a, p. 189). Since words are coded in two modes (text & picture or text & videos), they are learned more easily than those coded in one mode (text only). In addition, when words and phrases are instructed with different types of media, the retention is easier (Chun & Plass, 1996a). The rationale for the superiority of the dual mode of text and picture over text and video was that since participants could view the pictures for as long as they wished, the pictures allowed for the development of a mental model of the information; whereas, video annotations or clips are usually short; thus, left little room for establishing the information in long-term memory. In other words, as Chun & Plass (1996a) state, “the pace of the presentation of information is not sensitive to the cognitive constraints of the learner, and its transient character, therefore, does not allow the student to reflect and to refresh short-term memory” (p. 193).

Similarly, Al-Seghayer (2001) examined the impact of dynamic video glosses versus still-picture glosses among ESL learners in a multimedia learning environment. The study was guided by the question of which gloss combination, video clips or still-pictures, was more effective for facilitating vocabulary acquisition. Al-Seghayer (2001) hypothesized that video is a more effective tool to foster the acquisition of the new words in a foreign language. Thus, a software program was developed to provide students with three types of glosses (textual definition alone; textual definition and still- pictures; and textual definition and video clips). The glossed words were all in English. Al-Seghayer's (2001) study used a within-subject design in which the participants received the three modes of video annotations, text and picture annotation, and textual annotations. The findings of both recognition and production vocabulary tests revealed that words that were glossed with text and picture, and those with text and video were learned better than words with text-only glosses. Moreover, the participants recalled more words when the video clips were provided than when still-pictures or just text alone was present in short-term. Yet, the level of retention achieved and for how long (i.e., long-term retention) still needs further investigation. Al-Seghayer's (2001) study lent support to the dual coding theory of Paivio (1986), stating that when two verbal and visual materials (texts, pictures, and videos) are integrated, learners can meaningfully construct connections between two mental representations and can learn and retain the materials effectively. However, the findings of Al-Seghayer's (2001) research contradicted the results of Chun & Plass (1996a, 1996b), discussed above, whose studies showed that text and picture were more effective for recalling words than text and video glosses. The opposing findings, according to Cheng and Good (2009), may be attributed to some factors such as participants' different language proficiency level, the readability and appropriateness of the reading passages, the gloss types employed, the tasks used to demonstrate comprehension or recall, and the learners' language context. Other factors such as the type of assessment tools used to measure L2 learners' memory retention, the familiarity of the participants with the visual aids of pictures and video clips, the cultural background of the participants (i.e., types of participants attended the study), the target language, and the level of word difficulty (Mohsen & Balakumar, 2011) as well as learners' different learning styles (Plass et al., 1998) may also lead to the contradictory results for the two above-mentioned studies.

However, unlike the above two studies which revealed positive effects of multimedia glossing on L2 vocabulary acquisition, and supported the dual glossing modes over single mode, Akbulut (2007) carried out research resulting in no significant differences between the two visual modes of text and pictures and text and videos. There were three groups in his true experimental study design: two experimental groups that were instructed via the online hypertext annotations with the successive presentation of text and pictures or text and video clips respectively; and one control group which received the online text along with textual definitions and grammatical functions without access to picture or video clip annotations. All the participants were asked to indicate the words they could recall from the text (i.e., form recognition test) and write the L1/L2 equivalent of each target word they remembered (i.e., meaning production test). Also, the participants were requested to take a multiple-choice test, which had the components of the two previous tests (i.e., meaning recognition test) as their immediate and delayed vocabulary measurements. The findings indicated that the visual annotation groups (text

& pictures or text & videos) facilitated incidental vocabulary learning better than the verbal annotation group (text-definition alone); however, the two visual modes were equally effective. The findings here lend support to the generative theory of multimedia learning (Mayer, 1997) and confirm the positive and facilitative effects of multimedia glossing on L2 word learning and retention. The results suggest that presenting information verbally and visually is better than solely in words as they facilitate the "*transfer*" and "*retention*" of the information (Akbulut, 2007, p. 514). However, the conclusions are not in line with the results of the previous two studies, i.e., Chun and Plass (1996) and Al-Seghayer (2001) in regards to the annotation types that led to a better vocabulary learning and retention experience. The contradictory results might be attributed to participants' advanced level of language proficiency in Akbulut's (2007) study who did not benefit much from one visual annotation over the other. Other factors such as assessment methods (Chen, 2016), and types of picture or video assignments (Mohsen & Balakumar, 2011), might be determining criteria to change the findings. Therefore, further research is needed to address the controversies and inconsistencies regarding the effectiveness of different annotation types on L2 vocabulary learning and retention in multimedia-based setting with a focus on multimedia/hypertext glossing where the annotation types are presented simultaneously rather than successively.

Yoshii and Flaitz (2002) conducted a within-subject design study to investigate the effectiveness of the annotation types of text-only, picture-only, and a combination of the two on L2 incidental vocabulary retention in a multimedia reading setting. This study is the replication of Kost et al., (1999), who did an investigation in a non-multimedia environment and asked participants to read a story with pictorial, textual, and both pictorial and textual glosses. One hundred and fifty-one adult ESL learners were divided into three groups: the first group read a story with text annotation; the second group read the same story with picture annotation; and the third group received the same instruction with a combination of the two annotations (picture and text annotations). L2 learners' vocabulary gains were measured via immediate and delayed vocabulary post-tests. The findings, which were based on the three vocabulary instrument tests (i.e., picture recognition, word recognition, and definition supply tests), revealed that the combination group (annotations with text and picture) out-performed the text-only and picture-only groups across all measures (i.e., immediate and delayed post-tests). One interesting finding was that comparing the picture-only group with text-only group, it was found that the group did as well as the combination group not only on the picture recognition test, but also other measurements; whereas, the text-only group had significantly lower scores. One possible justification, as the researchers mentioned, would be due to the type of cues that the learners were exposed to (the participants of pictorial group received the words via pictures; thus, they had the chance to associate the meaning and word to the relevant images); while, the text-only group were not instructed through such cues. Yoshii and Flaitz (2002) suggest that "the picture cues were as effective as (or even slightly better than) the textual cues for the immediate retention of word meanings" (p. 45). Further investigation indicates that there was no significant interaction between annotation type and proficiency level for either the immediate or the delayed tests. In addition, the results confirmed the advantage of multimedia vocabulary glossing over traditional text glossing on enhancing vocabulary learning. The study again well supported Paivio's (1986) dual coding theory, which posits that presenting information is coded either verbally through texts and sounds or visually/non-verbally through pictures, images and video/animations (i.e., dual presentation) in mind. The researchers justified the effectiveness of dual glossing modes over single mode based on the interconnectedness of these two systems, meaning when words/texts are represented by one system (verbal), they can be activated by the other system (non-verbal) or vice-versa.

However, Yoshii and Flaitz's (2002) study was distinct from relevant studies, such as that of Kost and colleagues (1999), in that the participants here were first chosen from different countries speaking various languages. Thus, they were all multilingual; second, the study was conducted in a multimedia learning environment; and third, it engaged participants with two different levels of language proficiency (beginner and intermediate). However, the study could be replicated using a qualitative method of data analysis. Yoshii and Flaitz (2002) stated that "researchers could use a think-aloud protocol to observe what is happening and to speculate about cognitive processing when participants choose or not to look up the words" (p. 47). Additionally, different types of annotation/gloss types, like texts, pictures, audios, video and the combinations, could be examined to determine which one(s) are most effective in the combination. The study could be improved if longer exposure and multiple encounters with the words were included in order for FL/L2 learners to retrieve and retain the words faster and longer (Yoshii & Flaitz, 2002).

Employing a rather different gloss combination, Yeh and Wang (2003) investigated the effectiveness of text annotation only, text and picture annotations, and text and picture and sound annotations in association with FL learners' vocabulary learning in a multimedia setting. This study was different from others (Yoshii & Flaitz, 2002; Kost et al., 1999) as it used sound as a component of multimedia glosses, and that both first language (L1) and L2 were incorporated in textual glosses. The audio section of the instruction provided FL learners with a native speaker's voice, pronouncing a word, spelling the word, and reading aloud the sentence in which the word was embedded. The study also aimed at examining if learners' learning styles had any impact on different annotation types. Before the instruction, the participants took a vocabulary pre-test to measure their knowledge of the target words and a questionnaire, which asked their opinions about their perceptual learning styles. After the instruction, the learners were again given the vocabulary post-test, which included three tasks of word association questions, multiple choice questions on word meanings, and a cloze test. They were also requested to fill in the questionnaire regarding the design of the courseware followed by a short interview. The findings, which fall in line with the premises of dual coding theory and cognitive theory of multimedia learning, indicated that the version with text and picture was the most effective type of vocabulary annotation. The study also revealed that the aural information lowered vocabulary gains due to inadequate time allocation for the participants to complete the task, and the difficulty of the aural information which was presented at a speech rate that exceeded the participants' level of listening proficiency. Not being accustomed to use auditory skills in learning L2 words, the learners of this study found the sound annotation "*distracting*" rather than "*facilitating*" (Yeh & Wang, 2003, p. 140). Furthermore, the researchers justified the low vocabulary gain due to participants' different learning styles contributed to their low performance in aural annotation. Yeh and Wang (2003) report that the participants' preferred mode of instruction was visual/pictorial stimuli, since "they were not as strong when learning through auditory channels, as evidenced by the lower marks in auditory learning styles in the questionnaire" (p. 140); thus, their low skill in using phonological strategies was possibly one major reason why they did not perform well while learning L2 words through the combination of text, pictures, and audios. However, the results supported the hypothesis that texts and visual annotations are the most effective learning tools in multimedia learning environment. Yeh and Wang's (2003) study was also unique in that it offered FL learners both L1 and L2 texts, compared to other studies, which used either L1 text mode (Chun & Plass, 1996a; Kost et al., 1999; Plass et al., 1998) or L2 text mode (Yoshii & Flaitz, 2002; Al-Seghayer, 2001). Regarding the low performance of participants on the audio annotation mode, one possible interpretation is also the overloading of both visual and verbal information that might have excessed the cognitive processing. However, further investigation is required to know if the addition of audio glossing will impede or enhance L2 vocabulary learning and retention.

Drawing on the two prominent theories of multimedia learning[12], Mohsen (2011) conducted an experiment with two groups of participants (control & experimental) in order to investigate the impact of multimedia annotations/glossing on the acquisition of English language words. There were two group conditions: the experimental group who received the reading passage through multimedia software where the glossed words were annotated with texts, examples, images and audios; and the control group who were instructed with the same reading passages where the words were accompanied by definitions only (i.e., printed-text mode). Also, the participants of the multimedia group were required to fill in a questionnaire regarding their attitude towards the use of

[12] Paivio's (1986) dual coding theory & Mayer's (2001, 1997) generative theory of multimedia learning.

multimedia software. The findings of the two immediate and delayed vocabulary measurements (i.e., recognition & production) showed that the group who had access to multimedia glosses scored higher than those without any access to electronic annotations in both immediate and delayed post-tests, meaning the target words were learnt better when they were accompanied with texts, illustrated with examples, and annotated by audios and images than when they were learnt in sole (i.e., printed text or definition alone). Furthermore, there was a greater decline in the delayed post-test scores of the control group in comparison to that of the experimental group. The results lend support to the effectiveness of multimedia glossing on L2 vocabulary learning, and confirm the studies that used combinations of modalities such as texts, pictures, audios, and videos/animations. The use of varieties of media has also helped learners to retrieve the words better in long-term memory than when one mode was presented.

Adopting the three annotation modes of L1 definition alone, L1 definition and pictures, and L1 definition and video, Lin and Tseng (2012) conducted a quasi-experimental research study including both pre/immediate and delayed post-tests with a group of 88 beginner learners to investigate the effectiveness of dynamic video/animations and films versus pictures on L2 vocabulary learning. The researchers considered the application of video/animation as a facilitative tool in order to explain the difficult words, and present complex concepts (Weiss, Knowlton, & Morisson, 2002). Difficult words were defined as “the words that cause problems for the learners due to their unfamiliar spelling and concept” (Lin & Tseng, 2012, p. 348). The post-tests consisted of production and recognition vocabulary tests. The findings revealed that learning difficult words with textual definitions and videos was more effective than learning them with textual definitions and pictures or textual definitions alone. The reason could be justified based on the premise that presenting information through dynamic video/animation clips facilitated the construction of form-meaning connections for the difficult words and led the learners to easily build mental images of the target words; thus, enhancing the word retention in long-term (Lin & Tsneg, 2012). Besides, learners had the opportunity to experience meaningful learning as they were provided with a rich contextual and cultural “authenticity embedded in the target words” (p. 351). This rich context in the form of textual definition and video clips could help the learners stay focused and attentive to the new words, and recall the meanings of the words better for later use. On one hand, the results are aligned with Al-Seghayer’s (2001) study that showed L2 learners prefer to learn vocabulary with video clips rather than pictorial annotations. On the other hand, the findings are not in line with Chun and Plass (1996a, 1996b) who suggested that L2 learners favored learning vocabulary with pictorial annotations. However, other studies (Akbulut, 2007) found the two modalities of pictures and videos equal. Still, further studies should be conducted on the domain of multimedia vocabulary glossing to examine which gloss combination (textual, pictorial, aural and video/animation) would foster L2 vocabulary learning and promote long-term word retention.

In a pre/post-test study, Yanguas (2009) investigated the effectiveness of textual and pictorial multimedia glosses on text comprehension and vocabulary learning via a computerized text. The objective that guided the study was based on Schmidt’s (1990) theoretical framework of noticing hypothesis and attention, suggesting that learners’ attention would be attracted to word learning if the target words are prompted via a visual or verbal stimuli. Ninety-four participants were invited to read a computerized text under one of the four gloss conditions of textual, pictorial, textual and pictorial, and no gloss (control). Given the instruction, the participants were then asked to think aloud as they read the passages. Also, the researcher used an online protocol to track which gloss type was consulted/noticed more during the intervention. The purpose was to examine whether noticing would lead learners to better vocabulary learning. Two immediate and delayed recognition and production post-tests were used as assessment tasks during the study. Data from qualitative components revealed that most participants in the gloss groups noticed target words “at a low level of awareness” (Yanguas, 2009, p. 59). However, the quantitative components indicated that all the target glossed words were noticed and recognized significantly more by the multimedia gloss groups than the control group, and that no differential effects were found among any of the groups in the production of the target vocabulary items, meaning the appearance of annotation did not affect the production of word items (Bowles, 2004; Chun & Plass, 1996; Kost et al., 1999). In contrast to studies which argued that the appearance of glosses promoted vocabulary learning if the recall vocabulary test is used (Watanabe, 1997; Hulstijn et al., 1996), the insignificant difference in the production task for Yanguas’s (2009) study could be justified based on the fact that with productive vocabulary tasks, learners are not able to process all aspects of vocabulary learning such as morphology or lexical associations (Bowles, 2004; Lomicka, 1998). Furthermore, glosses were designed as a means of drawing learners’ attention to the target words and helping them comprehend the vocabulary items or texts better. Yanguas (2009) further explains that,

> glosses seemed to help participants make sense of the general meaning of the sentence but not make any type of lexical association that might signal deep

> processing of the word. This behavior by participants in the gloss groups seemed to be enough to perform significantly better than the control group in the recognition tests but not in the production tests. (Yanguas, 2009, p. 60)

However, the group with combination modes (i.e., textual & pictorial) outperformed the other groups in text comprehension and incidental vocabulary learning; and multimedia glossing resulted in a better vocabulary learning experience than either type in isolation. The findings were also in line with other studies suggesting the positive effectiveness of multimedia glossing on word learning (Al-Seghayer, 2001; Kost et al., 1999; Plass et al., 1998).

Salem and Aust (2007) conducted a mixed design study to examine the impact of different computerized text, audio, and picture glosses on reading comprehension, vocabulary acquisition and retention. Groups of intermediate ESL learners were randomly exposed to one of the five conditions of (a) no gloss; (b) gloss and English translation text; (c) text and spoken audio pronunciation in both Spanish and English; (d) text, audio and picture; and finally, (e) text, audio, picture and writing. English was the native language of all the participants except one Korean speaker. The results of the two immediate and delayed recognition and production vocabulary tests as well as the reading comprehension test showed that first, learners using glosses scored significantly higher than non-gloss users, and retained more new words in long-term. In simpler terms, those learners who used electronic glosses could understand Spanish words within the texts better, and thus recall them faster than those who did not have access to glosses of any type. Second, using glosses with text, audio and pictures was particularly beneficial in advancing the ability to write the meanings of glossed words in English in a productive vocabulary test, a difficult task comparing to the recognition vocabulary measurement. Additionally, language learners who had access to varieties of gloss types such as text and audios, text, audios and pictures, or text, audios, pictures and writing could obtain better reading comprehension scores, and acquired new words more than those with limited access to gloss types. The finding also signifies that learners who frequently consulted electronic glosses with several exposures to new words could achieve a better text comprehension and word acquisition. The results also aligned with Knight's (1994) findings that achieved a high correlation between the number of words students looked up (i.e., frequency of word) and their vocabulary test scores. The authors rationalized the outperformance of text, audio, and picture group in the two vocabulary tests of productive and recognition tests on the premise that enriching the glosses with well-designed pictures, clear textual definitions and native pronunciation can help establish "robust multi-sensory mental representations" (Salem & Aust, 2007, p. 5), suggesting that learners can acquire the new words more easily and retain them better in long-term if they have access to several gloss types (Clark & Paivio, 1991; Paivio, 1986). Also, the role of exposure is important in mediating between vocabulary acquisition and frequency of gloss application.

Aligned with the cognitive theory of multimedia learning (Mayer, 2014, 2005), Plass and colleagues (1998) conducted a research on a group of English speaking college students with the aim of investigating the effectiveness of multimedia glossing on individual learning differences as well as learning styles. The participants were asked to read a story in German presented via a computer program. The target words were available either through translation on the computer screen in English, verbal annotation with a native speaker pronouncing the word and its textual translation in English, picture and or video clip, visual annotation, or both. In taking the test, the participants had the option of producing the L1 translation of each new word and tick if the word reminded them of hearing it or seeing a picture or a video. The findings of this study revealed that students learned more effectively when they had access to visual and verbal annotation modes of material presentations than when only one or no mode was available. Also, given the choice of preferred mode of annotations, the participants could remember the L1 translations of the target words better than when they had no choice over the glossing modes. The results emphasized the importance of individual differences, such as learning preferences in the visualizer-verbalizer dichotomy for L2 word learning with media forms. Finally, Plass and colleagues (1998) concluded that words and pictures could best be acquired and recalled when presented dually through verbal and visual annotations. The bimodal presentation of information in Plass and colleagues (1998) signifies the relevance of the findings to the cognitive theory of multimedia learning (Mayer, 2014, 2005). The theory suggests that learners actively select relevant verbal and visual information, organize the information into the coherent mental representations, and integrate these newly constructed visual and verbal representations with one another.

However, Ariew and Erçetin's (2004) study showed the inconsistent results regarding the application of different gloss types in the form of hypermedia annotations[13] for intermediate and advanced ESL learners' reading comprehension. The researchers asked the participants to read an expository text using different forms of media such as textual definitions including text and audio, graphic and video annotations, and contextual definitions to provide information at the word level as well as topic level. The annotations were developed to help learners understand the text. The findings revealed

[13] "A hypertext presents information in nodes and links integrating textual information" (Ariew and Ercertin, 2004, pp. 237-238).

that the use of annotations did not facilitate reading comprehension; and there was a negative relationship between the amount of time on video/graphic annotations and reading comprehension for intermediate learners, suggesting that video was distracting for them; also, no relationship was found between the advanced learners' annotation use and reading comprehension, implying that their high linguistic competencies have enabled them to use good reading strategies to understand the text in the hypermedia mode. Also, learners' prior knowledge about the topic (i.e., the knowledge that was acquired before) could have had a role in helping them comprehend the text better. Nevertheless, the results of the questionnaire and interviews showed that the participants agreed on the usefulness of the annotations in the hypermedia environment; however, the textual definitions of words were more highly regarded more than words and audios because textual definitions "increased the speed of reading without getting bored" (Ariew & Ercetin, 2004, p. 253).

To support the efficacy of simultaneous multimedia glossing, Türk and Erçetin (2014) examined the impact of interactive versus simultaneous display condition of visual (text definition & picture) and verbal (text definition) glosses on incidental vocabulary learning and reading comprehension among eighty-two L2 low-proficient learners in a true experimental design study. In the interactive display mode, the participants were permitted to choose the type of multimedia information (visual or verbal), while in the simultaneous gloss condition; the information was presented in a single gloss mode at the same time. The participants were required to read an expository text; and their interactions and performance on the reading text were investigated through a recall protocol and a multiple-choice vocabulary test. Also, vocabulary learning was measured via several vocabulary tests of form recognition, meaning production, definition match, and bilingual synonym match. The findings indicated that the participants who were given control over their access/choice to the type of multimedia information (i.e., the interactive group) utilized glosses less frequently than those who were presented with both visual and verbal glosses simultaneously in a single gloss, implying that the simultaneous presentation of materials is an efficient and practical mode which allows learners to make use of “the text resources to a large extent” (Türk & Erçetin, 2014, p. 15) Furthermore, the simultaneous group performed significantly better on reading comprehension and vocabulary tests than the interactive group; although the time spent on the reading text was not significantly different between the two groups. In other words, simultaneous presentation of visual and verbal information leads to better vocabulary gains. The authors concluded that since the participants were among low proficient learners, their lack of language learning experience might have intervened with using texts with control over the type of multimedia information; thus, there is a need to train novice learners about “making strategic decisions in reading multimedia texts” (Türk & Erçetin, 2014, p. 17). The findings support one of the principles of generative theory of multimedia learning, namely called the contiguity principle (Mayer, 1997), positing the fact that when learners are presented with simultaneous verbal and visual information, they are able to build mental connections between the two representation modes in an integrated manner, which causes less cognitive load on their working memory (Mayer & Moreno, 2003; Mayer, 1997; Moreno & Mayer, 1999) and thus facilitates learning.

With the assumption of using animations as a tool to enhance L2 learners' comprehension and production of target words, Sato (2016) examined the effectiveness of computer-based visual glosses (pictorial vs. animated) on explicit L2 vocabulary learning of Japanese learners. The study drew on the cognitive linguistic framework of image schema, suggesting that "our concepts are meaningful because they are embodied or rooted in our bodily experiences" (Sato, 2016, p. 44). There were two groups (control & experimental) who received the three abstract spatial/locative prepositions of *above*, *on*, and *over* through pictorial and animated glossed formats. A preposition pre-test was conducted to gauge participants' knowledge of preposition use in appropriate context. At the end of the instruction, the two immediate and delayed receptive (i.e., fill in the blank) and productive (i.e., write English sentences using target words) vocabulary tests were also conducted to measure L2 learners' performance on the application of prepositions. The findings showed that there was no difference between the two groups in any test; however, animated glosses had positive impact on the text production than text comprehension. Thus, the use of images could facilitate L2 learners' sentence production with target prepositions.

Relying on form-focused instruction (Ellis, 2001; Long, 1991), and generative theory of multimedia learning (Mayer, 2001), Sadeghi, Khezrlou, and Modirkhameneh (2016) made a fine distinction between three vocabulary learning conditions (i.e., incidental, intentional, and explicit) and investigated the effectiveness of each condition on the vocabulary acquisition and reading comprehension of a group of upper-intermediate EFL learners by means of different hyper-text glosses in a multimedia learning environment. The gloss types consisted of text and pictures, text and audio, and text, picture, and audio. There were three experimental and one control groups. The participants completed Paribakht and Wesche's (1997) vocabulary knowledge scale (VKS) test, a contextualized vocabulary knowledge test and TOEFL as pre-tests to determine their unfamiliarity with the target words. They repeated VKS and contextualized test three months later to evaluate the long-term effect of the intervention. The findings of between-subject design showed that multimedia glosses positively influenced learners' vocabulary learning and reading comprehension; also, the text, picture, and audio gloss type outperformed the other gloss modes and resulted in better vocabulary learning and reading comprehension. However, the learning conditions affected participants' performance across gloss types in terms of both immediate and delayed post-tests. In other words, the glossing mode of text and picture was effective with incidental vocabulary instruction in the immediate contextualized vocabulary test; whereas, the participants in explicit and intentional groups achieved higher scores in the text, picture, and audio gloss type in VKS and reading comprehension tests. The results support Mayer's (2001) theory of multimedia learning and Paivio's (1990) dual assumption of dual-coding theory, suggesting that learners tend to learn better and effectively when both visual and verbal annotations are used rather than only one gloss type or no gloss mode; and Ellis' (2001) form-focused instruction that encourages attention to word forms and "teaches and uses words in non-communicative language tasks" (Sadeghi et al., 2016, p. 2).

In a recent multimedia-based research environment, Khezrlou and Ellis (2017) examined the impact of three learning conditions (i.e., explicit, incidental, and intentional learning) on FL learners' vocabulary acquisition and reading comprehension. Ninetynine participants received target word instructions in text, picture, and audio glossing modes (i.e., multi-glossing). The findings of the vocabulary tests of multiple-choice and written recall tests revealed that there were word gains for all the participants regardless of the exposed learning conditions; and explicit vocabulary instruction was the only condition that helped the learners to maintain target words in long-term. However, the reading comprehension test results varied according to the test types. In general, the study supported intentional and explicit instructional learning over the incidental learning condition. Besides, the use of multi-glossing facilitated the target word learning "even if the learners did not make deliberate attempts to learn the words" (Khezrlou & Ellis, 2017, p. 114).

Finally, Rassaei (2017) examined the effectiveness of textual (L1 definition) and audio glosses considering the effect of the two auditory and visual perceptual styles on L2 learners' vocabulary learning. The findings showed the efficacy of the two glossing modes over non-glossing strategy for word learning; besides, the results suggested that the participants who were categorized in auditory style of learning group and received the target words through audio glossing mode outperformed the visual learning group who were instructed through pictorial glossing mode. The findings also signify the prominent and determining role of learners' auditory-visual learning style and preferences for vocabulary learning. The superiority of audio mode for auditory group also supports the idea of matching the mode of instruction with perceptual learning styles.

Section C. Gaps in Literature and Relation to the Present Study

As mentioned in the previous section, the majority of the studies on multimedia glossing/annotations have contributed positively to the literature on promoting L2 vocabulary learning. Results of these studies confirm the claim that multiple glosses are more effective than a single or no gloss mode for learners' word learning (Al-Seghayer, 2001; Mohsen & Balakumar, 2011; Plass et al., 1998). Furthermore, when L2 learners are assigned to the combination of electronic gloss types with appropriate and clear textual definitions, relevant video/animations, and audible, well-recorded audios for the target words, they will be able to establish "robust multi-sensory mental representations" (Salem & Aust, 2007, p. 5), which can be a facilitating factor in accessing the words fast and retaining them longer for later use (Clark & Paivio, 1991; Paivio, 1991, 1986). However, research on multimedia/electronic glossing have shown inconclusive and inconsistent findings regarding which annotation/gloss type (s) (textual, aural, & visual) is more effective in facilitating L2 vocabulary learning and enhancing long-term word retention. Besides, less is written about the effectiveness of the multimedia gloss combination that incorporated L2 definition alone, L2 definition and audio, and L2 definition and video/animation on intentional vocabulary learning in terms of both short and long-term word recollection. Additionally, given the importance of gloss display mode on word retention (i.e., simultaneous or successive mode), the effect of simultaneous display condition of multimedia/electronic glossing on L2 vocabulary learning and long-term word retention has not yet been extensively focused. Finally, few studies have used mixed methods research to investigate multimedia glossing as an effective way to foster vocabulary learning and enhance word recollection in long term. Therefore, to address these limitations, I saw a need to open up another research area that investigated the effectiveness of simultaneous multimedia/electronic glossing on L2 vocabulary learning and word retention in short and long term. I pursued the present mixed methods research with the application of multimedia/electronic glossing in textual, aural and video/animation forms to examine which gloss combination better improved L2 learners' vocabulary learning for short and long-term use.

Chapter Summary

In this Chapter, I discussed the two-underlying theoretical frameworks of this study, namely Paivio's (1986) dual-coding theory and Mayer's (2014) cognitive theory of multimedia learning. I then presented a summary of literature on the two following themes: (a) vocabulary learning in language education, and (b) vocabulary glossing. I discussed multimedia glossing to promote vocabulary learning and long-term word retention, exploring the relevant literature on the domain. The gaps in the current literature, which prompted me to conduct this study, were then stated. In the next chapter, I describe the methodological approach I followed to address these gaps in the literature.

CHAPTER THREE: METHODOLOGY AND RESEARCH DESIGN

Introduction

The process of deciding on the appropriateness of a chosen methodology and its theoretical underpinnings is a vital component of accuracy in a research design (Appleton & King, 2002), which can also justify "the accomplishment of research aims" (Robey, 1996, p. 406). The present study aimed at investigating the effectiveness of simultaneous multimedia glossing on L2 learners' vocabulary learning and short and long-term word retention by providing them with a meaningful, contextual, and structured-based vocabulary teaching environment. This study followed a mixed methods research methodology[14] (MMR), and combined the elements of both quantitative and qualitative research approaches (Tashakkori & Teddlie, 1998). Since this research was neither purely empirical nor purely interpretive, employing mixed methods research offered a strong framework for the use of the quantitative and qualitative methods in order to develop a complete research analysis and to obtain a better and deeper understanding of the research problem than either approach alone (Archer, Bhaskar, Collier, Lawson, & Norrie, 1998; Danermark, Ekstrom, Jakobsen, & Karlsson, 2002; Tashakkori & Teddlie, 2003a, 2003b; Sayer, 2000; Venkatesh, Brown, & Bala, 2013).

The purpose of this chapter is to provide the justifications for the choice of mixed methods methodology as the research paradigm to conduct this investigation. The advantages and reasons for doing a mixed methods research are briefly explained with

[14] The terms mixed methods research methodology (Tashakkori & Teddlie, 1998) and mixed methods research (Creswell, 2003; Teddlie & Tashakkori, 2003) are used interchangeably in this chapter.

their relevance to the current study. The chapter is then followed by an overview of the types of methods and procedures employed to collect the data as well as the rationale for utilizing them. Also, the research design, participants, instrumentations, and procedures are fully discussed, each under a separate heading. Then, the procedures taken to gather the data are shortly described, along with some ethical considerations. Next, the general statistical procedures, taken to analyze the data, are explained and rationalized; and the homogeneity level of language learners are then checked and ensured. Finally, the chapter ends with a short summary.

Mixed Methods Research Methodology

Mixed Methods Research (MMR) is characterized by mixing qualitative and quantitative research methods, representing different research paradigms in the same research inquiry (Ågerfalk, 2013; Teddlie & Tashakkori, 2009; Venkatesh et al., 2013). It is sometimes referred to as the third methodological movement/paradigm (Teddlie & Tashakkori, 2009; Ridenour & Newman, 2008). Johnson and Onwuegbuzie (2004) define MMR as "the class of research where the researcher mixes or combines quantitative and or qualitative research techniques, methods, approaches, concepts, or language into a single study" (p. 17).

Mixed Methods Research Advantages and Rationale

There are two main benefits for the use of MMR over a single method approach in a study (Venkatesh et al., 2013; Teddlie & Tashakkori, 2009). First, employing MMR enables a researcher to develop a meta-inference[15], based on a combination of quantitative and qualitative data analyses (Danermark et al., 2002; Tashakkori & Teddlie, 2003a, 2003b; Venkatesh et al., 2013). Second, MMR provides an opportunity for a researcher to examine the research questions through one research method lens (e.g., quantitative research method) and complement the findings with another (e.g., qualitative research method) (Venkatesh et al., 2013; Teddlie & Tashakkori, 2009).

Also, researchers have shown that MMR is a highly recommended method of inquiry for the following reasons: *triangulation* (incorporating several methodological approaches, theoretical perspectives, data sources, viewpoints and methods of investigation to develop a deeper understanding of a phenomenon under study, and validate research findings); *complementarity* (using the findings from one method to clarify and illustrate the results from the other method); *initiation* (discovering contradictions and differences in research results that may lead to reframing research questions); *developmental* (using findings from one method to inform a research design involving another method); *expansion* (employing different methods for different inquiry components to expand the depth and breadth of the research); *diversity* (using different methods to combine researchers' and participants' perspectives through qualitative and quantitative research in order to identify different views of the same phenomenon); *completeness* (bringing together a more comprehensive account of the research area by employing both qualitative and quantitative research approaches in order to increase an in-depth understanding of the phenomenon and combine multiple methods and theories to discuss the results); and *confirmation* (using qualitative data to make new hypotheses,

[15] Meta-inference refers to developing an understanding of a phenomenon for which the use of either approach (qualitative or quantitative), in isolation, would be insufficient (Ågerfalk, 2013).

and employing quantitative research to test them within a single project) (Bryman, 1998; Creswell, 2003; Hussein, 2009; Johnson & Onwuegbuzie, 2004; Olsen, 2004; Tashakkori & Teddlie, 2008; Yeasmin & Rahman, 2012; Venkatesh et al., 2013).

Rationale to Use Mixed Methods Research

The reason I chose a mixed methods approach for this inquiry stemmed from my desire to produce findings which would be useful and valuable to the participants of this study, in particular, and larger English as a second language (ESL) population, in general. Additionally, mixed methods research coincided with one of the purposes of this research where few studies have been conducted using mixed methods to determine the effectiveness of the simultaneous multimedia glossing on L2 learners' vocabulary learning in regards to both short and long-term word retention. Therefore, employing multiple methods provided me with fresh insight and strong interpretations that were crucial to vocabulary learning and long-term word retention.

In general, I sought to employ mixed methods research in this doctoral study primarily for complementary and triangulation reasons. As for the complementary aspect, the findings from both the quantitative and qualitative components helped to illustrate how multimedia vocabulary glossing affected word learning of L2 learners and assisted them to enhance both short and long-term word retention. As for triangulation, since both research approaches focused on the same research questions and drew similar conclusions (Scott, 2010), I had a greater degree of confidence in the findings (Bryman, 1992). Besides, by using triangulation, I was able to increase the validity of the quantitative and qualitative findings, and improved both the internal consistency and generalizability of the results (Hussein, 2009; Tashakkori & Teddlie, 2008; Yeasmin &

Rahman, 2012), as well as the research instruments for measuring the effectiveness of multimedia glossing on L2 learners' vocabulary learning and word retention. Thus, the types of triangulation and their contribution to this present study are briefly discussed below.

Types of Triangulation

According to Denzin (1978, 1970) –as quoted in Angouri (2010); Bryman (1998), Hussein (2009), and Tashakkori and Teddlie (1998)– there are four different types of triangulation: (a) *data* triangulation, the use of several data sources to validate the results; (b) *investigator* triangulation, the involvement of more than one researcher to gather, analyse, and interpret the data; (c) *theoretical* triangulation, the use of multiple theoretical stances/positions in the study to interpret data and support and/or refute the findings; and (d) *methodological* triangulation, the use of more than one method of data collection and analysis (i.e., both quantitative and qualitative) in studying the same phenomenon. The methodological triangulation includes two types of between-method triangulation (combining and utilizing qualitative/quantitative data in order to obtain external validity) and within-method triangulation (using and crosschecking qualitative/quantitative data for internal consistency) (Yeasmin & Rahman, 2012). I achieved three of the four types of triangulation in this study: data, theoretical, and methodological triangulations.

Methodological triangulation in the present study.

From the methodological perspectives, I employed both between-and-within participant[16] designs to analyze the data and respond to the research questions. The following section discusses the two designs used in the study.

[16] Between-and-within participant/group designs are used interchangeably in this dissertation.

The between-participant design. This design compared the performance of the four groups of the participants in the study (groups A, B, C, and D)[17] to investigate if different glossing modes of instruction (i.e., L2 definition alone, L2 definition and audio glossing, and L2 definition and video/animation glossing) have any significant effect on L2 learners' vocabulary learning and their short and long-term word recollection.

The within-participant design. This design included the impact of the vocabulary modes of instruction (simultaneous multimedia glossing) across instructional sessions on L2 learners' vocabulary learning in terms of both short and long-term word retention. This study involved three instructional sessions clustered within a week (week 2) (See Table 3.1). The participants met the researcher every other day during one week (Saturday, Monday, and Wednesday) for the instructional sessions. As shown in Table 3.1, the time lag between the first instructional session to the next was one day (from Saturday to Monday or Monday to Wednesday) (i.e., short-term retention), and the interval between the final instructional session and the delayed post-tests was two weeks (14 days) (i.e., long-term retention).

[17] Groups are described in the procedure section of this chapter.

Table 3. 1 Week Intervals during Instructional Sessions

Week 2 (Instructional week)					Week 3 (Gap)					Week 4 (Delayed post-tests)				
Sat	Sun	Mon	Tue	Wed	Sat	Sun	Mon	Tue	Wed	Sat	Sun	Mon	Tue	Wed
					Two-week gap (14 days)									

End of instructional sessions/immediate post-tests

Delayed post-tests

(Sat = Saturday; Mon =Monday; Wed = Wednesday)

The rationale for a two-week gap stemmed from the work of Bahrick, Bahrick, Bahrick, and Bahrick (1993) who examined the retention interval (RI) of language learners' FL vocabulary acquisition and retention over nine years (long RIs). The findings of this longitudinal study revealed that increasing the retention interval from a fixed performance within each study session to 56 days resulted in better performance for the participants on the final tests (Bahrick et al., 1993). Due to the feasibility and access to the research participants, a two-week gap was considered as the retention interval in the present study.

Mixed Methods Research Key Decisions

Conducting mixed methods research is a decision that hinges on the research questions, problems, purpose, and context of the study (Venkatesh et al., 2013). Creswell and Clark (2011) outline four key decisions that are influential in the appropriate choice of mixed methods research design. They include:

1. Giving relative priority to the quantitative and/or qualitative data collection and analysis; this doctoral study had a greater emphasis on quantitative components than its qualitative counterpart because of the number of measurement tools and statistical analyses in the research; thus, the qualitative data only served as the supplementary/complementary to the findings in order to help the researcher gain a better understanding of the research outcomes.

2. Deciding on the appropriate stages at which qualitative and quantitative data is integrated and interacted; in this research, the data from the quantitative strand was first collected followed by the qualitative data, which were merely complementary (Creswell & Clark, 2011), and then mixed at the end of the study in order to infer conclusions during the interpretation and discussion stages (Creswell & Clark, 2011).

3. Implementing timing procedures; Creswell and Clark (2011) classify the order and timing in which the quantitative or qualitative data are gathered and analyzed within a mixed methods research design into three types: *concurrent*, "when the researcher implements both the quantitative and qualitative strands during a single phase of the research study;" *sequential*, "when the researcher implements the strands in two distinct phases, with the collection and analysis of one type of data occurring after the collection and analysis of the other type;" and *multiphase combination*, "when the researcher implements multiple phases that include sequential and/or concurrent timing over a program of study" (p. 66). The present research followed sequential procedures in that the empirical/quantitative phase of the data collection was conducted first, followed by the qualitative stage.

4. Using appropriate designs for mixing the qualitative and quantitative approaches; and finally,

5. Having an overall theoretical perspective to interpret the data (Creswell, 2003). I drew on Paivio's (1986) dual coding theory and Mayer's (2014, 2005) cognitive theory of multimedia learning to interpret the results of the inquiry.

In sum, the current quasi-experimental research gathered the data set in sequential procedures, while triangulating them by measuring both between and within-participant comparisons. The next section provides a detailed explanation regarding the implemented methods and research design in this project.

Methods

Research Design

Selecting an appropriate research design is pivotal in conducting MMR. Among the six possible categories for MMR designs (i.e., the convergent parallel design, the explanatory sequential design, the exploratory sequential design, the embedded design, the transformative design, and the multiphase design)[18], I adopted the embedded design, where I collected the quantitative data first, followed by the qualitative data to answer the research questions. The justification for choosing an embedded mixed methods design was that since the primary design of the research was experimental, I worked from the quantitative research methods first and used the qualitative method as "the supplemental method in service to the guiding approach" (Creswell & Clark, 2011, p. 92). In general,

[18]Review Creswell & Clark (2011) for the detailed discussion.

according to Tashakkori and Creswell's (2007) criteria, I followed QUAN$\rightarrow$(qual) design, where "QUAN" signified the priority of the quantitative data set over the qualitative components, The arrow "$\longrightarrow$" indicated the sequential order in the data collection, the bracket "( )" represented an embedded research design, and "qual" showed the secondary supportive (Greene, 2007) data set embedded into another. Together, the quantitative and qualitative findings from this research shed light on my understanding of the impact of simultaneous multimedia glossing on L2 learners' vocabulary learning in respect of short and long-term word retention. The next section describes the research participants, instruments, and procedures respectively. The justification of the choice of the statistical approaches is presented afterwards.

Research Participants

This study was conducted in a private language institute in Iran. In this language center, both English and French were taught and practiced by the students. I recruited 132 participants from intact classes in order to avoid changes in the main composition of each classroom and to not disturb their normal routine. In Iran, instructional classes are segregated—males and females are not usually placed in the same classroom. The participants were all young adults, ranging from 16 to 25 years old with an average age of 17.56. The participants' native language was Farsi (Iran's official language), and they learned English at the intermediate level. The participants attended the institute's language classes three times a week, and each instructional session lasted 90 minutes. After obtaining ethical approval from my institute, I gained access to the research participants in accordance with the private language institute's ethical policies in Iran. The participants' identities were concealed and protected through the use of pseudonyms.

This study assumed that the participants had similar linguistic and cultural background (Chen & Yen, 2013).

There were four groups in the study. The groups were labelled as group A (control group), and B, C, and D (the three experimental groups). Also, the experimental groups were labelled according to the order of gloss modes they received during the instructional sessions (See Table 3.2). Group B was characterized as TAV, where *T* stood for L2 definition alone, *A* referred to as L2 definition and audio glossing, and *V* represented L2 definition and video/animation glossing. This labelling also meant that group B received the target glossed words in the order of L2 definition alone, L2 definition and audio glossing, and L2 definition and video/animation glossing respectively (TAV). Similarly, group C was labelled as AVT; and group D as VTA. Group A was called the control group with no glossing instruction at all. Also, the participants in experimental groups received all the gloss modes in different orders.

Table 3. 2 Gloss-order Presentation

Groups	Gloss modes in three sessions
Gr. A	No glossing (control)
Gr. B	TAV
Gr. C	AVT
Gr. D	VTA

Proficiency assessment. I determined the proficiency level of the participants based upon three criteria: (a) language institute's placement test; (b) a Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001); and (c) two vocabulary pre-tests.

Language institute's placement test. The participants had already been assigned to the intermediate level before the study began based on the institute's placement test. The test included an audio and a written examination, as well as a face-to-face oral interview with the students. In the oral interview, two expert English language instructors at the institute asked questions from the Interchange Book (Interchange Book 2, 3rd edition, 2004), the main book of instruction at the institute, to determine learners' proficiency level. As an example, interchange student book level 2 is designed for language learners at an intermediate proficiency level. Thus, the language learners who successfully responded to most of the questions posed at this level belonged to the intermediate proficiency level and were included in the study. This test was administered two weeks prior to the beginning of the language term and was held in the language institute.

Vocabulary Levels Test. I used Schmitt and colleagues' (2001) Vocabulary Levels Test, version 2, at the 2000-word level (2k)[19] to assess participants' vocabulary knowledge and homogeneity, and to control for any proficiency differences among the participants attending the study (Peters, 2014).

Vocabulary pre-tests. I used the two vocabulary pre-tests of productive recall (PR) and multiple-choice (MC) productive recognition tests (See Appendices H & I respectively). The two tests served as the post-test instruments as well[20]. The reason to include them for the pre-test was to ascertain that the participants were unfamiliar with the target glossed words. The reason to administer them as the post-test was to determine the amount of vocabulary that the participants were able to retain both after the vocabulary instructions (immediate) and two weeks later (delayed). The PR test was given before the MC productive recognition test to avoid test effect and learning effect (Webb, 2005) for both pre-and post-tests. The two tests were administered in pencil-and-paper format. Timing for each pre-test completion was kept between 20-25 minutes. The

[19] See the instrumentation section for a detailed explanation of the Vocabulary Levels Test.
[20] Pre/post achievement tests are described in the instrumentation section.

binominal scoring of 1 (for each correct response) and 0 (for each incorrect response) was applied for the MC test; and the trinomial scoring of 2 (for each correct response), 1 (for partially correct response) and 0 (for each incorrect response) was employed for the PR test[21]. Also, according to Perez, Peters, Caltreboutm, and Desmet (2014), who set the 70% criterion for target word inclusion/retention, it was supposed that among the total 132 participants who attended the study, if 93 participants did not respond to the target glossed words correctly, those words would be kept for the test scoring. In other words, 30% of the participants (i.e., 39 people) should have known the target glossed words in order for the words to be excluded from the list. However, the results of the pre-test scoring showed that the highest number of the participants who did not know the target glossed words was even less than 30% of them (i.e., 39 individuals) in both PR and MC recognition pre-tests.

Research Instrumentation

The following instruments were utilized to gather the data: (a) demographic information form; (b) three reading comprehension passages adapted in three modes (hyperlinked with L2 definition alone, L2 definition and audio, as well as L2 definition and video/animation); (c) 33 target glossed words; (d) one Vocabulary Levels Test, (e) two immediate and delayed vocabulary measurement tests (PR and MC productive recognition tests); (f) a questionnaire; and (g) a semi-structured interview guide. I describe each instrument subsequently in detail.

Demographic information form.

Study participants were asked to fill in a demographic information form including their age, gender, length of time studying

[21] See vocabulary measurement section for the scoring rationale.

English, and how many other languages they knew in addition to English. The reason for collecting demographic information was to better understand the background of the participants when interpreting the data/findings (See Appendix C). The form took 5 minutes for the participants to complete. The analysis of the demographics of the participants showed that:

1. Group A consisted of male participants (n = 24); there were both male (n=26) and female (n=13) participants in group B; and the participants in groups C (n = 36) and D (n= 33) were all females.

2. There was only one participant in group C who knew 3 languages other than English; and all the other participants spoke Farsi as their native language and were learning English as their second language.

3. Only one participant from group C and one participant from group D had visited an English-speaking country; and the rest had never visited any English-speaking countries.

Tables 3.3 provides information about the demographic information of the participants in each group including number, gender, and the average of the age.

Table 3. 3 Demographic Information of the Participants including Number, Gender and Age Mean

Group	N*	Gender	Age mean
Gr. A (Control)	24	Male	17.33
Gr. B (TAV)	39	Male/Female	17.35
Gr. C (AVT)	36	Female	17.83
Gr. D (VTA)	33	Female	17.75

*There was no missing data in the sample.

Reading comprehension passages. Three expository[22] English reading passages were selected for the purpose of this study. The reading texts were: (a) *The Silk Road: Art and Archeology* (Hartman & Kirn, 2014, pp. 171-172); (b) *Problems in the Natural World* (Thaine, 2012, pp. 30-33); and (c) *Bites and Stings* (Zwier & Zimmerman, 2009, pp. 134-135) (See Appendices D, E, and F respectively for Text 1, 2 & 3). The texts were all adopted from intermediate-level English textbooks and were compatible with the proficiency level of the study participants. All the three texts were checked against the readability formulas and commented on by expert views.

As for the readability, various indices including Flesch-Kincaid grade level (Flesch, 1948), Gunning Fog score (Gunning, 1952), Coleman-Liau index (Coleman & Liau, 1975), SMOG index (McLaughlin, 1969), and the Automated Readability index (Senter & Smith, 1967) were calculated for the reading texts, and the three passages were at the readability level ranging from standard/average to fairly difficult to read, which was well-suited to the participants of the present study. Tables 3.4 and 3.5 show the text characteristics and the readability indices respectively.

Table 3. 4 Text Characteristics

	Text 1	**Text 2**	**Text 3**
Title	The Silk Road	Problem with the Natural World	Bites and Stings
Length (number of words)	337 words	399 words	389 words
Number of glossed words	12	11	10

[22] Passages that provide facts in a way that is educational and purposeful.

Table 3. 5 Readability Indices

	Text 1	Text 2	Text 3
Flesch Readability grade level	8.7	9.6	8
Gunning Fog score	10	12.4	10.4
The Coleman-Liau index	9	10	8
The SMOG index	7.6	9.3	7.6
Automated Readability index	8	9.7	7.7
Readability consensus	Standard/average	Fairly difficult to read	Standard/average

As for the expert views, three experienced ESL/EFL instructors, both from the same institute where this study was conducted and the language center where the pilot work was done, were asked to read the texts and comment if the passages were appropriate for ESL/EFL participants at the intermediate proficiency level. The three ESL/EFL instructors unanimously agreed that the passage contents fit the intermediate proficiency level, and that the topics were both unfamiliar and interesting to the learners. Therefore, I chose the reading texts as a means of vocabulary instruction in this study. There were four versions of the reading passages: (a) baseline version with no glossing, (b) L2 definition of the target words, (c) L2 definition of the target words with audio glossing, and (d) L2 definition of the target words with video/animation glossing. Each text was 300-400 words in length, and the word/lexical coverage was between 95 to 98% (Nation, 2013) (See Table 3.6 for word coverage before glossing the text); Read (2000) argues that "non-native speakers of English need to recognize at least 95% of the words in a text for efficient reading" (p. 83); whereas, Nation (2013, 1990) and Laufer (1997, 1992) contend that achieving at least 95 to 98% is necessary to meet this target. As the participants of this study were ESL learners, I took Nation and Laufer's range of 95 to 98% for the texts' word coverage.

Table 3. 6 Word Coverage before Glossing

	Text 1	Text 2	Text 3
Word Coverage	Before	Before	Before
	95.72%	95.72%	95.93%

Target glossed words. 33 unknown/target words (nouns) were selected to gloss in this study[23].

Vocabulary Levels Test (VLT). The test, originally designed by Nation (1983), was used "to supply a profile of learners' vocabulary, which is particularly useful for placement and diagnostic purposes" (Schmitt, 2010, p. 198). The rationale to employ VLT in this research stemmed from the findings of the studies, which revealed that "vocabulary size is directly related to the ability to use English in various ways" (Schmitt et al., 2001, p. 55). The test provides an accurate estimate of learners' vocabulary size at the targeted levels (Schmitt et al., 2001; Xing & Fulcher, 2007). Vocabulary Levels Test measures learners' knowledge of words at five levels: 2000, 3000, 5000, and 10,000 word levels and academic English words (See Schmitt et al., 2001; Nation, 1990). I utilized Schmitt and colleagues' (2001) VLT, version 2, at a 2000-word level (2k) due to its high validity and scalable profile of vocabulary frequency levels. The 2000-word level also provided the lexical resources for basic everyday oral communication, which was also well-suited to the intermediate proficiency level of the participants in this study. The 2000-word level contained 30 word questions in cluster (three noun clusters, two verb clusters, and one adjective cluster) (See Appendix G). As for scoring, Schmitt (2003, stated in Xing & Fulcher, 2007) recommends the score of 24 as the cutting point

[23] See word selection criteria in research procedure (phase 1).

for the acquired level, meaning that if the participants respond to 24 out of 30 (80%) questions correctly, they have acquired the level. Therefore, following Schmitt's (2003) scoring criterion, one full credit score was given for each correct response, and only those participants whose score was 24 or above for this vocabulary level (2k) were selected[24].

Vocabulary measurement tests. Research has shown that measuring vocabulary knowledge requires adopting a measurement method that "unambiguously measures just a single element of vocabulary knowledge" (Milton, 2009, p. 17). However, the two elements of reliability (the ability to test the vocabulary items consistently) and validity (whether the test item measures what it is supposed to measure) should be considered important in measuring vocabulary (Milton, 2008; Nation, 2013; Nation & Webb, 2011; Schmitt, 2010). Nation (2013) states that in order to measure vocabulary knowledge, one must consider "the purpose of the test, the kind of knowledge it will try to measure and the condition under which it will be used" (p. 514). As for the first option (test purpose), Nation (2013) refers to two vocabulary purposes that were also compatible with the objectives of the present study: (a) short-term achievement tests, where L2 learners were gauged to see if the studied group of words have been learned; and (b) long-term achievement tests, where the participants were measured to see if the instruction has been successful in teaching particular words.

In relation to the second criterion (i.e., types of word knowledge), Milton (2009) states that there are two categories for word knowledge: receptive/passive versus productive/active vocabulary knowledge; and breadth versus depth of word knowledge. Receptive/passive word knowledge, by definition, refers to the words that are known

[24] The homogeneity check based on VLT data is presented in the last section of this chapter.

when heard or read; whereas, learner's productive/active word knowledge entails the words that need to be recalled when one is using them in speech or writing (Milton, 2009). In the second distinction, breadth of word knowledge is the number of words a language learner knows; while the depth of word knowledge simply refers to the amount of knowledge a learner has about these words (Milton, 2009). Nation (1990) and Read (2000) also agree that receptive word knowledge refers to the ability of the language learners to recognize and recall the meaning of a word, while the productive word knowledge is counted for the ability of the learners to use the target words in speech or writing. This study followed Milton's (2009) receptive and productive classification of vocabulary knowledge, as well as Nation (1990) and Read's (2000) definition of receptive/productive word knowledge. Considering the fact that the participants of the present study were not taught the features of the target words, such as phonology, morphology, syntax and semantics (Laufer, 1990); or spelling, pronunciation, grammatical form, relative frequency and collocations (Nation, 1990), and that they only had limited exposure to the target glossed words during the instructional sessions, vocabulary tests were, thus, bounded into the category of receptive word knowledge, meaning the tests measured L2 learners' ability to recognize and recall the target glossed words from their meanings after the reading comprehension. Also, since the purpose of the study was to examine L2 learners' vocabulary learning, and not to measure their vocabulary knowledge, detailed discussion on measuring L2 vocabulary knowledge was beyond the scope of this dissertation, and thus avoided.

Additionally, following Nation and Webb (2011) who suggest that it is valid to measure learners' vocabulary learning by means of multiple tests, and that a single test cannot measure every aspect of word knowledge (Milton, 2009), the two vocabulary achievement instruments of PR vocabulary test and multiple-choice (MC) productive recognition vocabulary test were adopted. Webb (2005) states that employing both receptive and productive tests to gauge an aspect of word knowledge provides “a much more accurate assessment of the degree and type of learning that has occurred” (p. 50). The rationale for naming the two measurement tests as *productive* was that the participants had to recall the target words from memory (Cabeza, Kapur, Craik, McIntosh, Houle, & Tulving, 1977), and write them in the spaces provided (for PR test) or recognize the target words and choose them from the options given (for MC productive recognition) (Laufer & Goldstein, 2004; Nation, 2001).

Also, in regard to the *recognition* and *recall* category, Read (2000) defines *recognition* as when “test-takers are presented with the target word and are asked to show that they understand its meaning” (pp. 155-156); whereas, in *recall*, “they are provided with some stimulus designed to elicit the target word from their memory” (p. 155). However, Nation (2001) believes that “a recognition vocabulary test format involves the use of choices”; while, “a recall item requires the test-taker to provide the required form or meaning” (p. 359). To Nation (2001), MC productive recognition vocabulary tests “involve going from the meaning to the word form” (p. 359). Thus, following Nation’s (2001) category of a receptive recognition/productive recognition vocabulary test type, this study adopted the term MC productive recognition vocabulary format to give the meaning in the stem and require the recognition of the word form in the options provided. As for the PR test, this study took Read’s (2000) substitution of *recall* for the production vocabulary test and provided the meaning of the target words in the stem, asking for the recall of them in the space provided[25]. For the purpose of consistency and relevance, these two measurement tools are briefly described below in the same order as were given. Participants completed the PR test before the MC productive recognition test to avoid test effect.

Productive recall vocabulary pre/post-test. Production vocabulary tests are more challenging for FL/L2 learners compared to the recognition word tests (Nation, 2001). One common recall vocabulary test is definition-stem format, where the learners are asked to either provide the definition of a target word or recall it based on its definition in the stem (Öztürk, 2007). I designed the PR vocabulary test in the latter format, as shown below.

Example: 1. A mountain with a large opening which sends out burned materials:

___ ___ ___ ___ ___ ___ ___ (*Response: Volcano*)

The test items were designed based on the selected glossed words of the three reading passages. Reversed item-order was utilized to avoid guessing. The test contained 33 word-item questions (See Appendices H & M).

Multiple-choice productive recognition pre/post-test. Multiple-choice test items have been commonly used in standardized tests (Nation, 2013; Öztürk, 2007; Read, 2000) and among several glossing-based studies (Chen, 2016; Cheng & Good, 2009; Farvardin & Biria, 2011; Liu & Lin, 2011); and they typically measure two of the many taxonomies of word knowledge: the form and the meaning (Nation, 2001, 1990). If well-designed, MC tests can distinguish learners' partial vocabulary knowledge effectively (Nation, 2013; Read, 2000); and can be applied to a large number of learners in a short time (Öztürk,

[25] Two test samples (PR & MC productive recognition) are provided in the next section.

2007). However, multiple-choice items are fundamentally recognition tasks, where learners identify the correct response from among the alternatives (Cabeza et al., 1997). MC productive recognition tests can also be designed in contextualized (in a written context like a sentence) or decontextualized formats (out of context/in isolation) (Öztürk, 2007). I developed a MC productive recognition vocabulary test in a decontextualized format, where "the word was removed from its message context "(Nation, 2013, p. 103), and the study participants were required to choose the correct alternative among the four items (the correct item + 3 distractors) for each definition in the stem, as shown below.

Example:

1. A doctor who studies and treats skin diseases:

   a) Beautician b) Specialist c) Dermatologist d) Cosmetic surgeon

Response: item c

Similar to the PR test, a MC productive recognition vocabulary measurement tool was based on the 33-chosen target glossed words from the three reading passages with reversed item-order (See Appendices I & N).

Scoring criteria. The scoring criterion for the PR test was null (0) for wrong, blank, or incomprehensible responses, one full credit (1) for a partially correct response (such as minor misspellings or the substitution of one letter for another so long as it did not distort the meaning), and two full credits (2) for a completely correct response. The justification for this scoring criterion was based on Peters (2014), Türk and Erçetin (2014), and Al-Seghayer (2001). Peters (2014) scored all the post-test productive vocabulary tests dichotomously; the post-test was a productive vocabulary test in that a definition of a new word and its translation were given to the participants; the research participants were then required to write the new word. The first letter of the word was also written. A correct answer received one point, and an incorrect answer was given a zero point. The responses that contained “minor spelling mistakes such as ‘liabilitie’ instead of ‘liability’ were scored as correct" (Peters, 2014, p. 85). However, in the current study, the words such as “*bedbog*” for “*bedbug*”, “rudent” for “rodent”, and “*vose*” for “*vase*”[26] , to name just a few, were regarded as a sign of learning on behalf of the participants; and were thus counted as partially correct responses while scoring. Along the same vein, Türk and Erçetin (2014) assigned two points for each correct answer; one point for incomplete answers; and 0 points for wrong answers in the productive recall tests where participants were asked to supply the meaning of the target words in English or Turkish. Al-Seghayer (2001) considered 1 point for each correct or partially correct answer in a production vocabulary test. Also, similar to the pre-test, the scoring procedure for MC productive recognition test was full credit (1) for the correct response and null (0) for each incorrect response (Perez et al., 2014).

Each immediate post-test took 15-20 minutes for the participants to complete in pencil-and-paper format. The test items were piloted before the actual implementation, and were revised and reviewed by expert views for content appropriateness three times by two experienced language instructors (expert review). To achieve inter-item reliability for internal consistency, items were also checked for Cronbach alpha (α)[27].

[26] See Table 3.8 for a complete list of target glossed words.
[27] Cronbach's alpha is an estimate of internal consistency to measure how closely related a set of items are as a group.

The following section presents the findings of the reliability analysis on the two vocabulary measurement tests.

Reliability analysis of the measurement tools. Prior to analyzing the data, reliability coefficients were computed on the two pre/post achievement vocabulary tests of the PR and MC, using Cronbach alpha as a measure of the internal consistency. According to Table 3.7, the alphas were mostly above .7, which were acceptable at indicating the high internal consistency for the tests. Nunnally (1978), cited in Pallant (2013), recommends a minimum level of .7 for greater reliability.

Table 3. 7 Reliability Statistics for Pre/post MC and PR Vocabulary Tests

Test	Cronbach's Alpha	Number of Items
MC pre-test	.52	33
MC immediate post-test 1	.81	12
MC immediate post-test 2	.72	11
MC immediate post-test 3	.76	10
MC delayed post-test	.80	33
PR pre-test	.44	33
PR immediate post-test 1	.85	12
PR immediate post-test 2	.81	11
PR immediate post-test 3	.74	10
PR delayed post-test	.79	33

Nevertheless, the reliability coefficient for the PR pre-test ($\alpha = .4$) and the MC pre-test ($\alpha = .5$) were low. The reason for these low alpha levels could be attributed to the low variation of the scores, and the number of participants on the pre-tests. In other words, the majority of the participants did not perform well on the pre-tests, which resulted in low scores. However, on the subsequent administration of the tests (i.e., immediate and delayed post-tests), the participants performed better and the alpha level increased, which resulted in high level of variance among the test items. Besides, the two following criteria were considered while designing the test items: (a) not including the key words in the definition as it would make guessing easy; and (b) making target word definitions have similar lengths as the longer the definitions, the easier it would be to guess by making meaning associations.

Questionnaire. Questionnaires are one of the commonly used methods for collecting information (Colosi, 2006). I used both open-ended and close-ended questions in the present study. Open-ended questionnaires, as a type of survey, allow participants to provide responses and opinions with no restrictions (Colosi, 2006) and researcher's influence (Foddy, 1993). In contrast, close-ended questions limit the respondent to a set of pre-determined alternatives being offered (Reja, Manfreda, Hlebec, & Vehovar, 2003). However, there are two general advantages and one disadvantage for the use of an open-ended questionnaire. The advantages include: (a) noticing the varied responses that the participants give and their spontaneous answers (Reja et al., 2003); and (b) highlighting the ones that the researcher could not have anticipated (Colosi, 2006). The disadvantage is that the information received from an open-ended questionnaire may take extensive time to read and code in order to categorize the varied responses and identify the common ones (Colosi, 2006; Reja et al., 2003). I distributed a questionnaire[28] to assess participants' attitudes and perceptions towards the three glossing modes of vocabulary instruction (L2 definition alone, L2 definition and audio glossing, and L2 definition and video/animation glossing).

[28] See research procedure/phase 3 for the detailed discussion of the questionnaire.

Face-to-face semi-structured interview guide. Interviews are prominent data gathering methods in qualitative research as they access "people's perceptions, definitions of situations and constructions of reality" (Punch, 2011, p. 144). Interviews also have great flexibility to be used in a variety of research situations (Punch, 2011) and can yield in-depth responses about people's experiences, perceptions, opinions, feelings, and knowledge (Fontana & Frey, 2005; Patton, 2002). In general, interviews provide the opportunity for researchers to achieve crucial data, which cannot be acquired from observation alone (Gay, Mills, & Airasian, 2009; Patton, 2002). However, interview, in the present study, was mainly for complementary purposes to support the findings of the quantitative stage of data analysis.

Interviews can be categorized into different types. Patton (2002) classifies interviews as (a) informal conversational interview; (b) general interview guide approach; and (c) standardized open-ended interviews. In Fontana and Frey's (2005, 1994) category, interviews can take the form of structured, semi-structured, and unstructured types. As such, Minichiello, Aroni, Timewell, and Alexander (1990) provide a continuum of interviewing methods using the terms standardized, semi-standardized, and non-standardized interviews. This study followed Fontana and Frey's (2005, 1994) category of interviewing, and adopted a semi-structured interview guide to gather information about participants' reasons and preferences over the simultaneous presentation of the multimedia gloss types. A series of pre-determined questions were asked from the study participants with no limited set of response categories (Fontana &

Frey, 2005, 1994) in the semi-structured interview. The interview[29] sought attendees' opinions regarding which mode (s) of glossing they preferred and why.

Research Procedures

This study had three phases that included (a) the pilot study for selecting the target glossed words; (b) pre/post research procedures; and (c) questionnaire and interview research procedures. The description of each phase is presented below in consecutive order.

Phase 1. Pilot study: selecting the target glossed words.

Target glossed words were chosen from a pilot study, which was conducted both in Canada and Iran. In the piloting stage, 5-6 intermediate adult ESL/EFL participants ranging from 16 to 25 years old, were first given a consent/agreement form to sign (See Appendix A). The form ensured that the participants agreed to participate in the pilot study voluntarily. The participants were then invited to read three English reading passages and underline the words that they did not know. Each reading passage took 10-15 minutes of the participants' time. The purpose of the pilot study was to make sure that the participants understood the texts and underlined the words that they had little or no familiarity with. The selected target words were then used to instruct the research participants in the actual study.

The first pilot study took place in an institution with an English language center in Southwestern Ontario, Canada. The second pilot study took place in a small private language institute in Iran, other than the center where the actual study was conducted.

---

[29] See research procedure/phase 3 for a discussion on semi-structured interview.

The purpose of re-doing the pilot work with another 5-6 intermediate adult participants[30] in Iran was to ascertain/verify that the selected target words were also unfamiliar to those participants in that EFL context. All the participants at this pilot stage were not from the participants of the actual study. The participants' reading passages and responses were also kept anonymous. The results of the two pilot studies helped me to decide on the target words to gloss.

Criteria to select the target glossed words. I based the selection of 33 target glossed words on the following criteria:

a. The target words were selected based on the participants' unfamiliarity with the words. In order to achieve this criterion, in the two pilot studies, 10 adult intermediate ESL/EFL learners were invited to read three reading passages and underline the words they did not know. Words that were not familiar to at least 70% of the pilot participants (meaning, at least 3 participants in Canada and 3 in Iran) were chosen as the target glossed words (Perez et al., 2014).

b. Words were chosen that could be easily represented through simple and clear textual definitions and appropriate video/animation clips. I provided the L2 definitions for the target words from the Cambridge Advanced Learners' Dictionary (2013, 4th edition), the Oxford Learner's Dictionaries online, and Babylon version 7 (See word definitions in Appendix O). These three dictionaries offered concise, clear, and simple definitions for the words with several sentence examples. The definitions were edited

[30] The participants had the same age range as the target participants.

and reviewed by three English native speakers for the simplicity and length to ascertain that the word meanings were clear enough for all the participants to understand. The video/animation clips were selected from animation websites such as YouTube (www.youtube.com), Vimeo (www.vimeo.com), and Daily-Motion (www.dailymotion.com). The criteria were simplicity, clarity of word meaning (Al-Seghayer, 2016), and the quality of video/animation clips. In order to keep the timing consistent (between 7-10 seconds), and for the purpose of clarity and briefness, some sections of the video/animation clips were cut and edited. The animation/video clips were also piloted with three English native speakers. Changes were applied upon their suggestions. According to the lexical vocabulary profile, all the target word definitions were controlled for BNC - COCA - 25 (British National Corpus-corpus of Contemporary American English) to ensure that they were within 1000 to 2000 word levels (1k-2k). All the words used in the definitions were among the first 2000 word levels. The target glossed words were also checked for their word levels (WLs) according to the same lexical vocabulary profile (See Table 3.8). Of the 33 chosen glossed words, only one word (*goddess*) was in 1k; 5 words (*duster*, *feather*, *pitcher*, *ruins*, & *weapon*) were in 2k; 26 words ranged between 3k to 19k; and only one word (*bedbug*) was off the profile meaning that the word combination was among the very low frequency words; however, *bed* and *bug* were each in 1k and 2k lexical profile respectively.

c. Not too many words were glossed in each passage. Studies on multimedia glossing have shown that depending on the text-word counts, the number of glossed words differ (Al-Seghayer, 2001; Bell & Le Blanc, 2000; Chun & Plass, 1996a, 1996b; Farvardin & Biria, 2012; Ko, 2012; Lin & Tseng, 2012; Türk & Erçetin, 2014; Yoshii, 2006). As an example, Farvardin and Biria (2012) glossed 30 words in a 901 word-text; whereas, Al-Seghayer (2001) used a 1300 word-text with 25 target words to gloss. Yoshii (2006) highlighted 14 glossed words in a 390 word-text, but Bell and Le Blanc (2000) glossed 67 words with a 409 word-text. In a recent study, Türk and Erçetin (2014) adapted a 980 word-passage with 35 glossed words. Besides, the findings of Choi's (2016) study revealed that glossing too many words may have "detrimental" effects in ways to cause L2 learners to focus excessively on glossed words without trying "to extract correct word meanings from contexts" through guessing or contextual clues (p. 143). These experimental studies reveal that around 2 to 5% of the words would need to be glossed in a passage (Nation, 2013). Nation (2013) contends that in order to adapt a text with glossed words, two criteria should be taken into account: vocabulary frequency and density of unknown words. In relation to density data of the text, he states that "the highest density of glossing should be no more than 5%, and preferably around 3%, of the running words" (p. 242). Regarding vocabulary frequency, he asserts that it is feasible "to gloss mid-frequency words, and replace low-frequency words" (Nation, 2013, p. 242), because low-frequency words contain only 1% of learners' word coverage (the 10th 1000 words), whereas mid-frequency words contain 95% of learners' word coverage (the 3rd 1000 words to 9th 1000 words) (See Nation, 2013, pp. 14-19). With that said, if FL/L2 learners know 3000-4000 words of English (the 95%-coverage), in a passage of 300-word page, between 3.53 to 5.79 % of words should be glossed (3.53% means 10 words per 300-word page, and 5.79 % means 17 words per 300-word page) (See Nation, 2013, p. 243). Now, if we assume that there are ten words in each line in a passage, only one word for every two to five lines needs to be glossed (Nation, 2013), which matches the 2 to 5% gloss coverage in a text above. Thus, following Nation's gloss percentage, three reading passages were adapted, with each passage containing 300-400 words. Twelve words were glossed in passage 1, eleven words in passage 2, and ten words in passage 3.

d. The words were selected according to the degree of importance, interest, and usefulness for understanding the reading passages (Hong, 2010; Ko, 2005; Nation, 2013, 1990). The criterion was achieved by asking three ESL/EFL teachers to read the words in the reading passage according to their importance to the plot of the story (Elley, 1989 as cited in Nation, 2013). Thus, words appearing more frequently in a reading passage had a higher coverage level, and were regarded as important words (Nation, 2013, 2001).

e. Concrete noun-words were chosen to gloss in this study as they were easier to provide both L2 definitions and video/animations than the verbs, adjectives or adverbs. According to Mohsen and Balakumar (2011), concrete words can be imaged by different visual modes because “they are tangible to the senses” (p.153). Choosing noun words to gloss also controlled for the part of speech (Jung, 2016; Laufer & Rozovski-Roitblat, 2011), since they are the most frequently occurring part of speech in a reading passage (Webb, 2005).

f. Contrary to Brown, Waring, and Donkaewbua (2008) and Jung (2015) who controlled for the word-frequency variable in their study, this study did not control the word frequency, because the hypothesis was that multiple exposures to the glossed words could help L2 learners to retain them over a long time. Yet, some of the selected target words, though few, appeared more than once in the reading passages of this research, ranging from twice to three times (See Table 3.8). Besides, research has shown that increasing the number of times unknown words are encountered in a text increases the potential and possibility for vocabulary learning and acquisition (Chen & Truscott, 2010; Eckerth & Tavakoli, 2012; Horst, Cobb, & Meara, 1998; Laufer & Rozovski-Roitblat, 2011; Peters, 2014; Rott, 1999; Webb, 2010, 2007; Waring & Takaki, 2003). According to Schmitt’s (2008) claim, “anything that leads to more exposure, attention, manipulation, or time spent on lexical items" can add to learners' vocabulary learning (p. 340). Additionally, the amount of time the learners interact and engage with the words facilitate vocabulary learning (Yusuf, Sim, & Su'ad, 2014; Schmitt, 2010).

g. Finally, the target glossed words were annotated with L2 definition alone, L2 definition and audio glossing, and L2 definitions and video/animation glossing. All target words were bold-faced (Roby, 1999) in dark blue, underlined, and hyperlinked. This hyperlink or textual enhancement signified that a gloss for the word was available, and that it provided the opportunity for learners' mental retrieval (Nation, 2013) and had a positive impact on vocabulary acquisition (De Ridder, 2002). Table 3.8 shows the list of the 33 target glossed words used in this study as well as their word levels (WLs) and frequency of occurrence in the text.

Table 3. 8 List of Glossed Words with Word Level and Frequency of Occurrence

	Target glossed words/ WL/Frequency of occurrence		Target glossed words/ WL/Frequency of occurrence		Target glossed words/ WL/Frequency of occurrence
1	Archeologist (K-3) (Once)	12	Duster (K-2) (Twice)	23	Mosquito (K-5) (Once)
2	Arachnid (K-17) (Once)	13	Feather (K-2) (Once)	24	Orchard (K-6) (Once)
3	Armor (K-4) (Once)	14	Flea (K-7) (Once)	25	Pesticide (K-5) (Once)
4	Bedbug (off) (Once)	15	Goddess (K-1) (Once)	26	Pitcher (K-2) (Twice)
5	Beetle (K-6) (Once)	16	Hip (K-3) (Once)	27	Pollinator (K-8) (Three times)
6	Bumblebee (K-14) (Once)	17	Horsefly (K-19) (Once)	28	Rodent (K-7) (Once)
7	Crop (K-3) (Twice)	18	Ivory (K-5) (Once)	29	Ruins (K-2) (Once)
8	Competitor (K-3) (Once)	19	Ladder (K-4) (Once)	30	Tomb (K-5) (Once)
9	Deer (K-4) (Twice)	20	Linen (K-5) (Once)	31	Vase (K-5) (Once)
10	Dermatologist (K-9) (Once)	21	Merchant (K-4) (Once)	32	Venom (K-7) (Once)
11	Volcano (K-4) (Once)	22	Wasp (K-6) (Once)	33	Weapon (K-2) (Once)

Phase 2. Pre/post-test research procedures. The research took place in an English language institute in Iran and included the 8 following stages: The first 4 stages occurred a week before the actual vocabulary instruction (pre-instructional session/week 1); the next 2 stages occurred in week 2 (instructional sessions); and the last 2 stages happened in weeks 3 and 4 respectively (post-instructional sessions). The detailed description of each phase/week follows:

Pre-instructional session (week 1).

1. One week prior to the vocabulary instruction, one hundred and thirty-two adult EFL participants at the intermediate level were given a consent/agreement form to sign (See Appendix B). The form ensured that the participants agreed to participate in the actual study voluntarily. The consent form included all the details regarding who conducted the study, what the intent/purpose of the study was, and the potential benefits or risks that might come from the study. The consent form also outlined the confidentiality procedures and specified the participants' rights in the study. The estimated time to complete the form was 3-5 minutes. The participants were told in advance that the study would take four weeks, and that they would receive small gifts as incentives if they took part in all the study sessions. The aim was first, to appreciate their participation and time, and second, to encourage them to attend the instructional sessions and do the tests for a full data analysis.

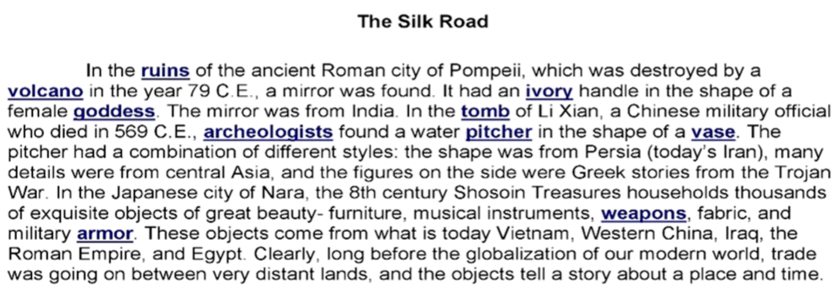
2. The participants were then assigned to 4 groups of A (control), and B (TAV), C (AVT) and D (VTA) (experimental groups) from their intact classes and were given name tags to wear on the days of the instruction (i.e., participants assigned to group A wore "*Group **A***" name tag as did the other groups). The number of the participants in groups ranged between 24 to 39 language learners (Table 3.9) with the same intermediate proficiency level. Group A received no glossing instruction; whereas all the participants in groups B, C and D received the target glossed words through the three glossing modes of L2 definition alone, L2 definition and audio glossing, and L2 definition and video/animation glossing.

3. After signing the consent form and filling out the demographic information form, the participants were then given a Vocabulary Levels Test (VLT) to fill. The test functioned as a pre-test to gauge participants' proficiency level, and took 15-22 minutes for them to complete.

4. As soon as the participants completed the VLT and handed in the test sheets, they were given two other vocabulary pre-tests (PR and MC productive recognition tests).

Instructional sessions (week 2).

5. The vocabulary instruction took place in one week during three consecutive sessions, every other day. The participants completed the three sessions in a large classroom equipped with a computer, two speakers, a monitor, and two big blackboards. I delivered the words in order to eliminate the effect of different teaching styles by another teacher. During the instruction on each day, the participants were referred to the reading passage of that day shown on a large screen monitor in front of them—the glossed words were hyperlinked and boldfaced. Ten to Twelve glossed words were taught in each session. The instruction time for each word was kept between 5-7 seconds. Each instructional session took 20 minutes, and the total length of the instruction was 60 minutes for both the control and three experimental groups. In relation to the *"how"* of the instruction, in week 2 (the week of instruction), the participants in each experimental group (B, C, & D) received the target glossed words through all the glossing modes of instruction (L2 definition alone, L2 and audio glossing, and L2 and video/animation glossing) (See Table 3.10, Week 2), whereas group A (control) received no specific interventions and external enhancement. For example, in week 2, day 1, the participants in **group A** received *text 1* with no specific instruction of vocabulary glossing; the instructor only read aloud the passage from the large screen monitor, making a pause equal to 5-7 seconds after each glossed word. Participants in **group B** received the same text (*text 1*) via L2 definition alone. While reading the text aloud, the instructor clicked on each target glossed word, which was in bold, underlined, and hyperlinked on the screen. A window popped up showing the glossed word with its simple definition. The participants in **group C** received the same text (*text 1*) through L2 definition and audio glossing. Clicking on each word, a window was opened that showed both the definition of the target glossed word and its audio recording. The audio file was recorded by the voice of an English native speaker and was controlled for timing (5-7 seconds). Finally, the participants in **group D** received the target glossed words in *text 1* via L2 definition and video/animation glossing. This time, when a word was clicked on, a window was displayed with the definition and a short video/animation clip of that word (approximately 7-10 seconds). The modes of instruction were counterbalanced for each group in the following/subsequent days as the major purpose of the study was to investigate the effectiveness of different glossing modes on vocabulary learning. Thus, the texts remained the same for each group on each day to eliminate the text-effect on learners' performance at the end of the instruction; and the glossing modes were the only items altered. Furthermore, all of the participants were given a hard copy of the texts with no target word underlined (See Figures 3.1-3.4 for the screenshots of the multimedia instruments).


The Silk Road

In the **ruins** of the ancient Roman city of Pompeii, which was destroyed by a **volcano** in the year 79 C.E., a mirror was found. It had an **ivory** handle in the shape of a female **goddess**. The mirror was from India. In the **tomb** of Li Xian, a Chinese military official who died in 569 C.E., **archeologists** found a water **pitcher** in the shape of a **vase**. The pitcher had a combination of different styles: the shape was from Persia (today's Iran), many details were from central Asia, and the figures on the side were Greek stories from the Trojan War. In the Japanese city of Nara, the 8th century Shosoin Treasures households thousands of exquisite objects of great beauty- furniture, musical instruments, **weapons**, fabric, and military **armor**. These objects come from what is today Vietnam, Western China, Iraq, the Roman Empire, and Egypt. Clearly, long before the globalization of our modern world, trade was going on between very distant lands, and the objects tell a story about a place and time.

Figure 3. 1 Sample text for Group A (Control)

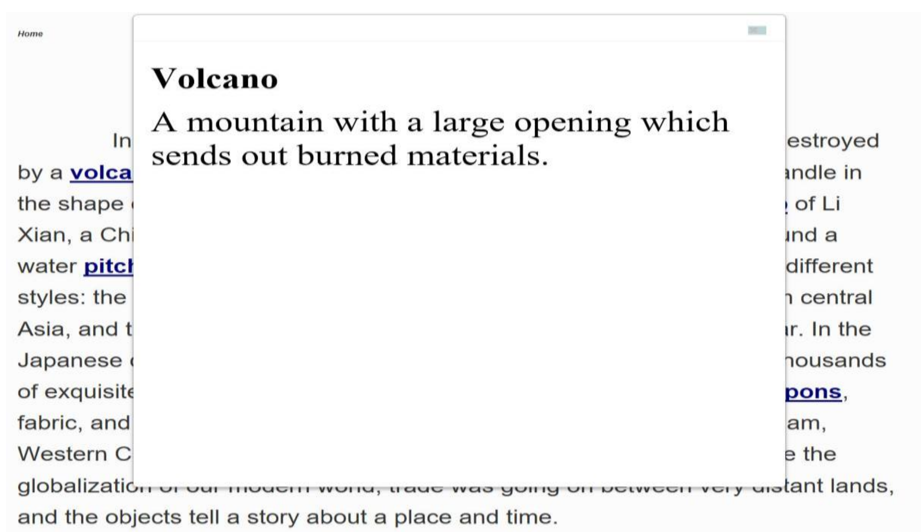


Figure 3. 2 Sample text for text/L2 definition alone (single glossing)

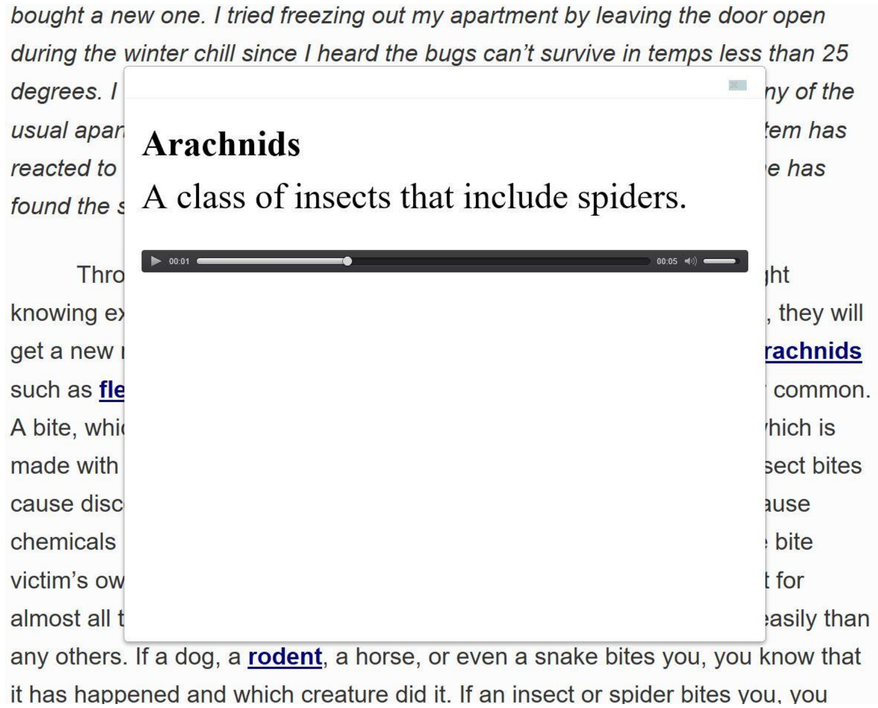


Figure 3. 3 Sample text for text/L2 definition & audio glossing (dual glossing)

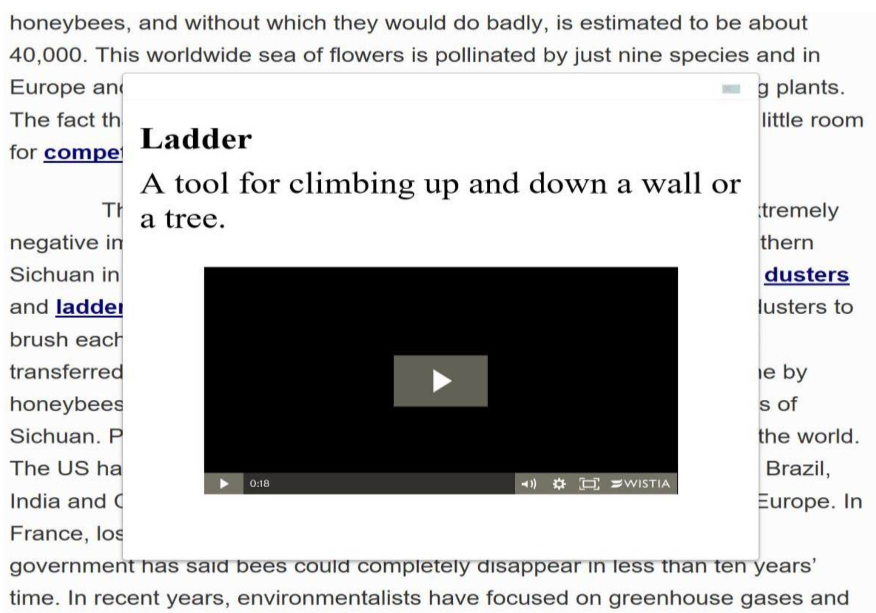


Figure 3. 4 Sample text for text/L2 definition & video/animation glossing (dual glossing)

6. After the instruction on each day, the participants of all groups were given two immediate vocabulary post-tests (PR test and MC productive recognition test) from the same text of that day (See Appendices J, K & L). The purpose was to measure participants' short-term word retention. The same procedure was also followed for day 2 and day 3, with text 2 and text 3. At the end of the third instructional session, I asked the participants to meet again two weeks (14 days) later. There was no activity or instruction in week 3.

Post-instructional sessions (weeks 4 & 5).

7. On a scheduled day two weeks after the instruction (week 4), the participants were asked to do the two delayed post-tests (PR test & MC productive recognition test). The purpose was to measure participants' long-term word retention.

8. On the same day, after the delayed post-tests, a questionnaire was distributed to the participants of the experimental groups, which took 10-15 minutes for them to complete. The participants were then rewarded for their time and full participation throughout the study. After completing the questionnaire, a certificate, prepared by the language institute, and a small gift, were given to the participants as a token of the researcher's appreciation. The participants were also asked to indicate their preference and availability for an interview for the following week (week 5). Those who attended the interview sessions were also given another small gift for their full contribution and commitment. Table 3.9 presents the number of participants attending each session, and Table 3.10 illustrates the detailed description of the research procedure.

Table 3. 9 Number of Participants in Each Session

Week sessions / Groups	Week 1 (pre-instruction)	Week 2 (Instruction)			Week 3 (Gap)	Week 4 (post-instruction)	Week 5 (interview)
		Session 1	Session 2	Session 3			
Gr. A (Control)	# 24	#24	#24	#22	----	#20	----
Gr. B (TAV)	# 39	#30	#30	#31	----	#28	#3
Gr. C (AVT)	#36	#33	#30	#30	----	#27	#3
Gr. D (VTA)	#33	#32	#31	#31	----	#28	#3
Total	132	119	115	114	----	103	9

Table 3. 10 Overview of the Research Procedure

<table>
<tr>
<th rowspan="2">Pre-instructional session (week 1)</th>
<th colspan="7">Instructional sessions (week 2)</th>
<th rowspan="2">Week 3

(Gap)</th>
<th rowspan="2">Post-instructional session

(Week 4)</th>
<th rowspan="2">Post-instructional session

(Week 5)</th>
</tr>
<tr>
<td>Groups</td>
<td colspan="2">Day 1

(Session 1)</td>
<td colspan="2">Day 2

(Session 2)</td>
<td colspan="2">Day 3

(Session 3)</td>
</tr>
<tr>
<td rowspan="4">• Consent form
• Demographic information form
• Pre-test (1): (Vocabulary Levels Test)
• Pre-tests (2): (PR & MC productive recognition tests)
• Participants' assignment to groups: (A, B, C & D)</td>
<td>A</td>
<td colspan="6">Control group (no glossing)</td>
<td rowspan="4">There was no instruction or activity during this week</td>
<td rowspan="4">2 delayed post-tests (productive recall & MC productive recognition tests)

&

Questionnaire</td>
<td rowspan="4">Face-to-face semi-structured interviews</td>
</tr>
<tr>
<td>B</td>
<td>Text 1

L2 definition alone</td>
<td rowspan="3">2 immediate post-tests from text 1</td>
<td>Text 2

L2 definition and audio</td>
<td rowspan="3">2 immediate post-tests from text 2</td>
<td>Text 3

L2 definition and video/ animation</td>
<td rowspan="3">2 immediate post-tests from text 3</td>
</tr>
<tr>
<td>C</td>
<td>Text 1

L2 definition and audio</td>
<td>Text 2

L2 definition and video/ animation</td>
<td>Text 3

L2 definition alone</td>
</tr>
<tr>
<td>D</td>
<td>Text 1

L2 definition and video/animation</td>
<td>Text 2

L2 definition alone</td>
<td>Text 3

L2 definition and audio</td>
</tr>
</table>

Phase 3. Questionnaire and interview research procedures. The qualitative study procedure contained 2 stages that follow:

9. As stated above, a questionnaire was distributed to the participants of groups B, C, and D (the experimental groups) after the two delayed post-tests. The questionnaire included 15 open and close-ended questions (i.e., thirteen rating questions and two open-ended questions) (See Appendix P). As for the rating questions, the participants were asked to respond to the first nine questions using a 5-point Likert-type scale, with 1 being strongly agree and 5 being strongly disagree (i.e. 1= strongly agree; 2= agree; 3 = neither agree nor disagree; 4 = disagree; 5 = strongly disagree) (questions 1-9). The participants were also asked to rate the degree of helpfulness of the vocabulary instructions in the next three 5-point Likert questions, with 1 being extremely helpful and 5 being unhelpful (i.e., 1= extremely helpful; 2=helpful; 3= somewhat helpful; 4= neither helpful nor unhelpful; 5=unhelpful) (questions 10-12). The participants were also asked to rank their preferences(s) towards multimedia glossing modes from 1 to 3, with 1 being their first priority and 3 their least preference (question 13). In addition to the previous questions, the two-other follow-up open-ended questions sought participants' opinions upon the type(s) of glossing mode they received that assisted them to learn and remember the target glossed words easier as well as their rationale (question 14). Also, the participants were asked to express their additional comments regarding other vocabulary instructional strategies and techniques they have used that facilitated their word learning (question 15). Learners were required to indicate their preference and availability for an interview for the following week as well (week 5). The questionnaire was given in English; however, only a few participants asked for clarification questions on some parts of the questionnaire, which I addressed them individually in their first language (Farsi) for a full understanding.

10. In week 5, I interviewed the participants upon their willingness to take part in the face-to-face semi-structured interview. The purpose of the interview was to know which mode(s) of glossing (L2 definition alone, L2 definition and audio glossing, or L2 definition and video/animation glossing) the participants preferred and why. The priority for participant selection was given to those who completely attended all the study phases including pre-test, instructional and post-test sessions. Among the 25 participants who volunteered, nine people were selected for the interview (3 people from each experimental group). Also, 3 participants from the control group were interviewed to seek their opinions about how their vocabulary learning would be influenced/or be different if instructed via the modes of L2 definition alone, L2 definition and audio glossing, and or L2 definition and video/animation glossing. The aim was to know more about their vocabulary learning strategies and preferences; therefore, the questions were different in that they did not specifically ask about the three modes of vocabulary glossing. This is why the data on this section was not included in this dissertation. An interview guide, which was self-developed and piloted twice, was used to structure the interview (See Appendix Q). Even though the interview allowed the interviewees to venture into a conversation beyond the questions, it was essential that I adhered to the predetermined list of the questions to discover common or conflicting themes between participants. The interview was conducted in the language institute, and the length was kept approximately 30 minutes for each participant. The data from the interviews were also audio-recorded, transcribed, and reviewed phrase by phrase; and the frequent or dominant themes inherent in statements were then extracted.

Data Collection and Ethical Consideration

Once I entered the language classroom, I collected the consent form and the demographic information form as well as the three pre-tests (a Vocabulary Levels Test and two vocabulary pre-tests). After each instructional session, I collected the results of the two immediate post-tests (PR & MC productive recognition tests). I also gathered the results of the delayed post-tests two weeks after the instruction. After the delayed post-tests, a questionnaire was administered; the relevant data were collected, the responses were coded, and common themes were identified.

Six months prior to conducting the research, I submitted the study outline entailing all the details of the project to my institute's research ethics board (REB), and followed the ethical guidelines and principles of the institution where this PhD work was being pursued (See Appendices S & T).

General Data Analysis Procedures (Quantitative Component)

From the quantitative perspective, data analysis procedures consisted of employing univariate (i.e., descriptive statistics) that included means and standard deviations for each scale and any subscales that emerged through principal component analysis. SPSS 20.0 software (IBM Corp., 2011) was used to analyze the quantitative data. The level of significance was set at an alpha level of $p < .05$. For the analyses where the statistical test assumptions[31] were not met, the alpha level was set at $p < .025$ (Tabachnick & Fidelle, 2013). Relevant statistical data were presented with tables and graphical figures. From the qualitative perspective, this study examined L2 learners' attitudes and perceptions towards glossing modes of vocabulary instruction via a questionnaire and face-to-face semi-structured interviews. The purpose of the next section is to introduce the overall statistical procedures used to analyze the quantitative data. The rationale to conduct the statistical testing methods are also discussed.

As stated earlier in this chapter, two types of between and within-participant comparisons were carried out to investigate the research questions; and the data from two immediate and delayed pre/post-achievement tests were collected. Furthermore, the number of the test items as well as the test contents were the same for both the pre-test and the delayed post-test; whereas each immediate post-test was a sub-test of both the pre-test and delayed post-test. It was supposed that comparing the vocabulary scores of the immediate post-tests to those of the pre-tests would show participants' short-term word retention; comparing the vocabulary scores of the delayed post-tests to those of the immediate post-tests would display the extent to which target glossed words were retained from short-term to long-term; and comparing the vocabulary scores of the pre-test to those of the delayed post-tests would reveal participants' long-term word retention.

[31] Assumptions are discussed in the next section.

I conducted a 4 (four groups in the study) ×2 (two immediate post-tests) ×2 (two delayed post-tests) analysis of covariance (ANCOVA) for all the between-participant comparisons in this study in order to investigate the impact of glossing and glossing modes on L2 learners' vocabulary learning with respect to short and long-term word retention. Also, paired samples t-tests were utilized for all the within-participant comparisons to examine the effectiveness of glossing and glossing modes on L2 learners' vocabulary learning in regards to short and long-term word retention.

Variables

Independent variables (IVs).

The primary IV in this study included different glossing modes of instructions/groups for between-participant design. The glossing modes were L2 definition alone, L2 definition and audio glossing, and L2 definition and video/animation glossing; and the secondary IV was the test sessions for within-participant design. Other variables such as text-effect and test effect were controlled. Text effect was controlled by keeping the same passage constant for each group on each instructional day; and test effect was taken into account by administering the PR test before the MC productive recognition vocabulary test.

Dependent variables (DVs).

All test scores, including immediate and delayed post-test scores were considered as the DVs in both ANCOVA and paired samples t-test.

The following section discusses the rationale to conduct the statistical testing methods of ANCOVA and paired samples t-test, used in this research. The normality of all the data scores has also been checked[32], and the preliminary assumption testing was conducted for both ANCOVA and paired samples t-test. All the statistical analyses were

[32] See the next section for the discussion of normality.

done twice: once for the PR vocabulary test and once for the MC productive recognition vocabulary test.

ANCOVA

Analysis of covariance, namely referred to as ANCOVA, is an extension of the analysis of variance (ANOVA) (Tabachnik & Fidell, 2007), and is generally used when there is one dependent variable to examine differences between groups/independent variables while statistically controlling for the effect of another additional variable (Pallant, 2013), called a covariate (CV). The covariates are the continuous variables "that are not part of the main experimental manipulation", but have an influence on the dependent variable(s) (Field, 2009, p. 396). Controlling for the effect of a covariate would reduce error variance[33] and assure the researcher if other variables in the study were not confounding the observed outcome (Mayer, 2013; Pallant, 2013). In other words, by using ANCOVA, a researcher is able to identify the significant mean differences among the groups (Pallant, 2013) after "adjusting the mean of DVs on one covariate" (Tabachnik & Fidell, 2007, p. 245). In fact, adjusted means are defined as "the means that would have occurred if all the subjects had the same scores on the covariates" (Tabachnick & Fidelle, 2013, p. 200).

There are mainly two purposes to use ANCOVA in a study: (a) ANCOVA increases the power of the test for controlling the main effects (Pallant, 2013; Tabachnik & Fidell, 2007); and (b) it adjusts the mean scores on the DV, considering the covariate effects. ANCOVA is common in research settings where the participants are taken from intact classes or cannot be randomly assigned to the intervention conditions, but are

[33] Using the covariate reduces the possibility of Type II error (Marefat et al., 2016).

recruited from the existing groups (Pallant, 2013; Tabachnik & Fidell, 2007). In occasions like the intact classes, such as the recruitment condition in the present study and with the existence of pre/post-testing design, ANCOVA controls for all the pre-test score differences (Larsen-Hall, 2010), so that "the only differences that remain are related to the effects of the groupings (IVs)" (Tabachnik & Fidell, 2007, p. 196).

Rationale to use ANCOVA. The rationale to use ANCOVA in this study had to do with controlling the initial pre-test differences while comparing the immediate and delayed post-test scores across all the 4 groups (control and the three experimental groups). In the present study, since the number of the participants in each group was not the same in the pre-test, and it was assumed that the groups might have performed differently from one another, these initial pre-test differences were controlled via the ANCOVA test. In other words, removing them would increase the chance of detecting any differences among the groups (Pallant, 2013). Also, as stated earlier, the sub-tests of both pre-test and delayed post-test were created; therefore, in all ANCOVA tests, the 3 immediate post-tests, and the one delayed post-test with three sub-components acted as DVs; the groups/different gloss modes acted as one IV; and the initial pre-test scores were considered as a covariate. The immediate post-tests were independent from one another, but their preceding instructions were on the same set of the words that were used in both the pre-test and the corresponding delayed post-test. In other words, each pre-test consisted of 33 target word-items which were administered in one single session before the instruction began (week 1); each immediate post-test included a sub-component of the pre-test and the delayed post-test, and were administered in 3 separate sessions immediately after the instructional sessions (week 2). The immediate post-test 1 contained 12 word items; the immediate post-test 2 contained 11 word items; and the immediate post-test 3 consisted of 10 word items. Finally, the delayed post-test with 33 word-items was administered two weeks after the instruction (week 4). This is why the immediate post-tests were compared separately each time considering the initial pre-test effects. ANCOVA also aimed to investigate if the glossing technique, in general, and different glossing modes, in particular, have had any significant effect on L2 learners' vocabulary learning in regards to short and long-term word retention. Several pairwise comparisons were also done to examine where the significant differences lay among the four groups.

The assumptions of ANCOVA. Conducting ANCOVA as a parametric test required the

fulfilment of some of the main assumptions pertinent to this analysis. The important statistical

assumptions that were controlled in this study were: (a) normality of data; (b) Levene's Test (1960) for homogeneity of variances; and (c) Homogeneity of regression slopes (Field, 2009; Pallant, 2013; Tabachnik & Fidell, 2007). Each assumption is briefly explained below.

(a) Normality of data. This assumption checks if the population from which the samples are taken are normally distributed around the means (Martin & Bridgmon, 2012; Field, 2009). Normality is often checked to decide between running a parametric test (e.g., one-way ANOVA)[34] or its non-parametric equivalence (e.g., Kruskal Wallis Test)[35] for the data analysis. To check normality, two procedures are often followed: (a) looking at the values of Skewness and Kurtosis[36] in the SPSS output (i.e., descriptive tables); and/or (b) comparing the scores in the sample to a normally distributed set of scores with the same mean and standard deviation via the Kolmogorov–Smirnov or Shapiro–Wilk tests (Field, 2009). In the present study, the normality of the data was assumed through both Skewness and Kurtosis ratios, and the Kolmogorov–Smirnov test. In the former case, if the Skewness and Kurtosis ratios fitted between the range of ± 1.96, the data was normal (Pallant, 2013; Field, 2009). This assumption was met for most of the data. Only a very few data were not normal, which could be overlooked for two reasons: (a) the sample/group sizes from which the data was collected were not different from each other; and thus, if very small portion of the data is not normal, the violation will not affect the rest of the data (Field, 2009); and (b) ANCOVA, as a robust statistical analysis against violation of normality, was employed to run the analysis (Field, 2013). In the latter case (i.e., the Kolmogorov–Smirnov test), if p value is greater than .05 ($p > .05$), then the data is normal (Field, 2009). This test showed that, in the present study, except for the VLT data set, all the data were normal for the pre/post and delayed tests. Thus, to analyze the results of VLT scores, a non-parametric test was used.

---

[34] One-way ANOVA is used for significant differences between two or more means when the data are normal (Martin & Bridgmon, 2012)

[35] Kruskal-Wallis Test is used when the data are not normal. This test makes the comparison of three or more groups possible by converting mean to mean ranks for each group, and relying on the ranks for the analysis (Pallant, 2013).

[36]Kurtosis is an index of peakedness of the distribution, and Skewness is an index of symmetry of the distribution. If a sample is normally distributed, its skewness ratio and kurtosis ratios (i.e. skewness and kurtosis divided by their standard error) should not go beyond ± 1.96 (Pallant, 2013). In occasions where Kolmogorov-Smirnov test is not available or the samples are too large, Skewness and Kurtosis ratios are employed (Tabachnick & Fidell, 2007).

(b) Levene's Test for homogeneity of variance. This assumption was checked via Levene's Test (Levene, 1960) to make sure if groups were homogenous, meaning there were equal variances among the group in a distribution (Pallant, 2013), and throughout the data especially "in designs with several groups of participants" (Field, 2009, p. 150). In other words, each of these samples should come from populations with the same variance. Levene's Test of homogeneity of variance was applied for between-participant comparisons in this study. For the homogeneity of variance to be assumed, the p value should be greater than .05 ($p > .05$) (Pallant, 2013; Field, 2009). This alpha level was met when checking Levene's Test throughout the data analysis; however, in occasions where the assumption of homogeneity of variances was not perceived, "a more conservative alpha level" was set to determine the significance for that variable (Pallant, 2013, p. 304). Tabachnick and Fidell (2013) suggest an alpha of .025 rather than the conventional .05. Therefore, in the main ANCOVA results, any time this assumption was violated, the alpha for significant differences was set at .025.

(c) Homogeneity of regression slopes. The assumption checked if the relationship/interaction between the covariate and dependent variable was the same for each of the groups. Pallant (2013) states that "unequal slopes would indicate that there is an interaction between the covariate and the treatment" (p. 310). Also, it is assumed that if p value is greater than .05 ($p > .05$), the assumption is perceived (Pallant, 2013). This assumption has been controlled in the third rows of all the main ANCOVA tables throughout this study.

Paired samples t-test

A t-test is a univariate statistical testing method used to compare the mean scores for the same group of people in a study on two different occasions and timing intervals such as the pre/post-testing conditions (Pallant, 20103). T-test determines whether or not groups are different while comparing their mean scores (Field, 2009). There are generally two types of t-tests: (a) independent-sample t-test, and (b) paired samples t-test, also referred to as the repeated measures (RMs) (Pallant, 2013). The former is used when there are two experimental conditions where different groups of participants are assigned to each; whereas, the latter is employed when the same participants in one group are assigned to the two experimental conditions in pair at different time intervals (Field, 2013; Pallant, 2013); so "the two mean scores cannot be independent of each other" (Field, 2009, p. 138). A t-test design has one DV and at least one IV. In situations where there is pre/post and delayed test scores, like the one in the current study, a paired t-test (repeated measurement) is commonly used for within-participant comparisons (Larson-Hall, 2010).

Rationale to use paired samples t-test.

The rationale to employ paired samples t-test analysis had to do with examining if glossing and different gloss modes were effective for L2 learners' vocabulary learning in terms of long-term word retention (i.e., the analysis showed the extent to which words were retained in the long-term from short-term). Thus, the performance of each group was compared from the immediate post-tests to the delayed post-test. Since each group was compared on two different testing occasions (immediate and delayed), the paired samples t-test was applied as an appropriate technique of analysis for the within-participant comparison in this research.

The assumption of paired samples t-test. Normality of the data, as one of the main assumptions of paired samples t-test (Pallant, 2013), was ensured throughout the data analysis.

General Data Analysis Procedures (Qualitative Component)

From the qualitative perspective, I used a questionnaire, as a survey, as well as semi-structured interviews to collect the data. As for the questionnaire, from the total of 132 participants, only 83 learners attended the session in which the questionnaire was administered (i.e., week 4 of the study); 49 participants were absent on that day. The survey responses were then analyzed to examine if learners' attitudes and perceptions regarding word learning and retention had been influenced through different glossing mode(s)[37]. As for the interviews, 9 participants volunteered to be interviewed (week 5). I audio-recorded the interviewees' voices for further data analysis process. All interviews were transcribed in full. Regarding the transcriptions, I used regular Window's Media Player and Word software to document the transcriptions. After transcribing the data, I reviewed them to ascertain that the statements were accurate. To code the data, I followed manual coding, and extracted the themes that recursively were occurring throughout the document. I underlined the most frequent themes, and kept them for further analysis (inductive coding) (Brundrett & Rhodes, 2014; Miles, Huberman, & Saldaña, 2014; Miles & Huberman, 1994). I sub-categorized the themes which assisted me to find the themes that related to the research questions and further clarified the findings of the quantitative part of this study[38].

[37] The analyses of the questionnaire are described in detail in chapter *Four (Result)* of this dissertation.
[38] The themes as well as relevant analyses are described in chapter *Four (Result)* of this dissertation.

The next section presents the procedure taken to check the homogeneity level of the participants in VLT scores.

Homogeneity Check: Vocabulary Levels Test

As stated earlier, I used a Vocabulary Levels Test (VLT), version 2 (Schmitt et al., 2001) to ensure the homogeneity of the participants in terms of their vocabulary knowledge at the beginning of the study. After the VLT test was scored, all 132 participants were qualified to take part in the study. The normality of the VLT data was checked (See Appendix R – Table R.1). The result showed that the data were not normal. Therefore, Kruskal Wallis Test was used to compare the means of the four groups (control & three experimental groups) to make sure that the participants in the four groups were not significantly different from each other; thus, homogenous. Table 3.11 presents the descriptive statistics of the VLT. As shown, the mean scores of the 4 groups are relatively similar, despite the slight low performance of the participants in groups A (control) and B (TAV) in comparison to groups C (AVT) and D (VTA).

Table 3. 11 Descriptive Statistics: Results of VLT pre-test

Group	N	Mean	Std. Deviation	Skewness		Kurtosis	
					Std. Error		Std. Error
Gr. A (Control)	24	23.45	3.93	- 2.72	.47	8.20	.91
Gr. B (TAV)	39	23.38	3.39	-2.73	.37	10.48	.74
Gr. C (AVT)	36	24.41	1.91	-.86	.39	1.57	.76
Gr. D (VTA)	33	24.51	1.90	-.74	.40	.58	.79

Note: T = L2 definition alone; A = L2 definition & audio glossing; V = L2 definition & Video/animation glossing.

Table 3.12 changes the mean score to mean rank for each group; and Table 3.13 presents the main findings of Kruskal-Wallis Test. According to the table, the four groups were not significantly different from one another in terms of the VLT mean scores; $\chi^2 = 2.71$, $df = 3$, $p = .43$ ($p > .05$)[39]. Therefore, it can be concluded that the control group (A) and the three experimental groups (B, C, & D) were homogenous at the beginning of the study with respect to the Vocabulary Levels Test.

Table 3. 12 Kruskal-Wallis Test (Changing Mean Scores to Mean Rank for VLT Data)

	Groups	N	Mean Rank
VLT Pre-test	Gr. A (Control)	24	64.94
	Gr. B (TAV)	39	59.12
	Gr. C (AVT)	36	70.08
	Gr. D (VTA)	33	72.45

Table 3. 13 Test Statistics[a,b] (Kruskal-Wallis Test Result)

	VLT Pre-test
Chi-Square	2.71
Df	3
Asymp. Sig.	.438

a. Kruskal Wallis Test
b. Grouping Variable: Group

Summary

This study aimed to investigate the efficacy of simultaneous multimedia glossing on L2 learners' vocabulary learning in terms of both short and long-term word retention. In this chapter, I discussed the methodological aspects of the study including the research design, participants, instrumentation and procedures. The data collection and data analysis were also explained, followed by a review of the general statistical procedures adopted to address the research questions. The rationale to employ each statistical method was discussed, and the homogeneity of the participants was ensured.

[39] In Kruskal-Wallis test, if p value is smaller than .05 (p <.05), groups are significantly different from one another; in contrast, if p value is larger than .05 (p > .05), groups are not significantly different; thus homogenous (Pallant, 2013).

The two statistical methods of ANCOVA and paired samples t-test were utilized to measure the between-and-within participant comparisons. The data were collected through pre/post achievement tests immediately and two-week after the intervention. L2 learners' attitudes and perceptions towards multimedia glossing modes of vocabulary instruction were examined through a questionnaire and face-to-face semi-structured interview.

The next chapter presents the findings of this study in terms of the two vocabulary tests of PR and MC productive recognition tests respectively as well as questionnaire and semi-structured interview.

CHAPTER FOUR: RESULTS AND ANALYSIS

Introduction

The purpose of this study was to investigate the efficacy of simultaneous multimedia glossing on L2 learners' vocabulary learning with respect to short and long-term word retention. This chapter presents the analysis of the results based on the two vocabulary measurement tools of productive recall (PR) and multiple-choice (MC) productive recognition tests as well as the questionnaire and semi-structured interview to address the research questions and sub-questions. The chapter is divided into three sections: section one presents the data analysis on the PR vocabulary test; section two analyzes the data from the MC productive recognition vocabulary test; and section three interprets the findings of the questionnaire and face-to-face semi-structured interview.

As mentioned in the previous chapter, all the quantitative calculations were done on the pre/post immediate and delayed test scores. Also, for the purpose of the comparisons, the sub pre-tests as well as the sub immediate and sub delayed post-test scores were calculated/created; and the analyses were done first with in-sum comparisons for the impact of glossing, and another time with sub-test comparisons for the impact of different glossing modes. In-sum test comparison considered the total scores of the participants in pre/post and delayed tests; whereas, sub-test comparisons took into account participants' scores on each individual test. A brief summary of the findings is then presented at the end of each section in regard to the research questions.

SECTION ONE

(Productive Recall Test)

In-Sum Comparisons

The in-sum comparisons address the following research question:

1. *Do different glossing modes (L2 definition alone, L2 definition and audio glossing or L2 definition and video/animation glossing) have any significant impact on L2 learners' vocabulary learning and short and long-term word retention?*

For the purpose of clarity and ease of follow, this research question is divided into two sub-questions (1.1 & 1.3); however, for the purpose of the conciseness, another analysis was done to compare participants' scores from immediate to delayed post-tests for short versus long-term effect (i.e., 1.2). The following section addresses questions *1 (1)* to *1 (3)*.

Pre-test in-sum to Immediate Post-test in-sum (Between-participant Comparison)

1 (1) *Does glossing have any significant impact on L2 learners' vocabulary learning for short-term word retention?*

In order to investigate the impact of glossing on L2 learners' vocabulary learning in terms of short-term word retention, the performance of the four groups (control & three experimental groups) on the pre-tests in-sum was compared to the sum of the immediate post-tests (i.e., between-group comparison). Figure 4.1 as well as Table 4.1 show the trend of changes in the vocabulary mean scores of each group across the test sessions from the pre-test in-sum to the delayed post-test in-sum. The means of each group are out of 33 in Table 4.1.

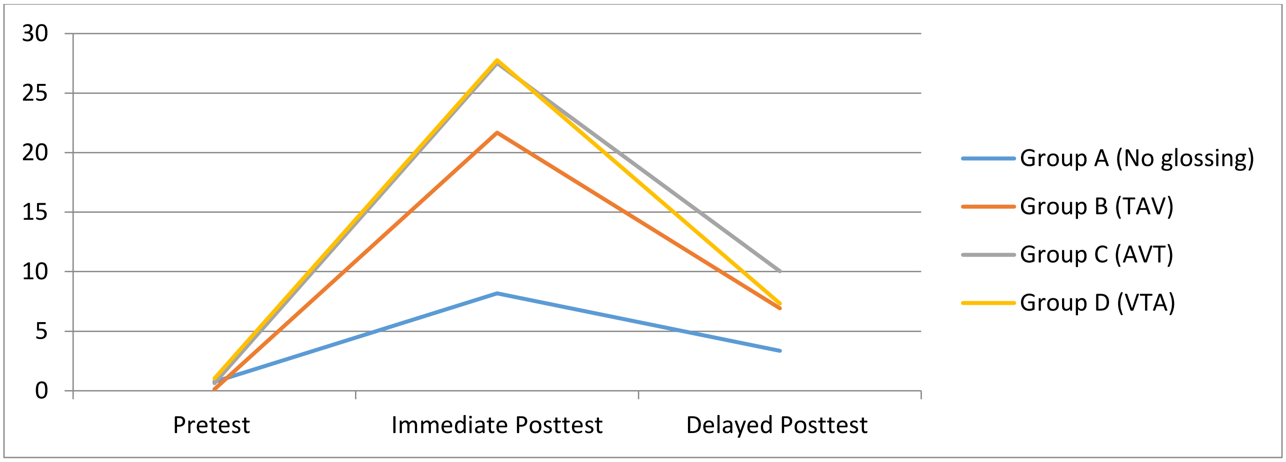


Figure 4. 1 Trend of changes in groups across sessions from pre-test to delayed post-test in sum (PR test)

Table 4. 1 Descriptive Statistics of the 4 Groups from Pre-test to Delayed Post-test in-sum (PR test)

Group		N	Mean	Std. deviation	Skewness		Kurtosis	
						Std. error		Std. error
Gr. A (Control)	Pretest	24	.75	1.32	1.599	.472	1.326	.918
	Immediate post-test	22	8.18	3.64	.216	.491	-1.319	.953
	Delayed post-test	20	3.35	3.15	.416	.512	-.921	.992
Gr. B (TAV)	Pre-test	39	.12	.46	3.684	.378	12.759	.741
	Immediate post-test	28	21.67	11.23	1.016	.441	.923	.858
	Delayed post-test	28	6.92	4.79	1.457	.441	3.218	.858
Gr. C (AVT)	Pre-test	36	.63	1.45	3.024	.393	9.269	.768
	Immediate post-test	27	27.51	8.67	.069	.448	.239	.872
	Delayed post-test	27	10.03	8.30	1.380	.448	3.688	.872
Gr. D (VTA)	Pre-test	33	1.03	1.87	2.361	.409	6.012	.798
	Immediate post-test	30	27.76	9.66	.220	.427	-.509	.833
	Delayed post-test	28	7.32	6.22	.492	.441	-1.175	.858

As the means of groups A, B, C, and D in Table 4.1 show, the performance of the four groups on the pre-test was very low, showing that the groups had a minimum baseline knowledge regarding the target glossed words before the instruction; however, with instruction, the three experimental groups (B, C, & D) who received the glossed words via different glossing modes performed better than group A (control), because they received higher mean scores compared to the control group. Furthermore, group A (control) had the lowest mean score (M=8.18) on the immediate post-tests in comparison to the three experimental groups. Among the three experimental groups, it was group D (VTA) who had the highest mean score (M=27.76) on the immediate post-tests in-sum; group C (AVT) had the second highest mean score (M=27.51); and group B (TAV) received the lowest mean score (M=21.67). Also, the descriptive table and the visual graph unanimously depict that the performance of the four groups (A, B, C, & D) declined from the immediate post-tests to the delayed post-test considering the groups' vocabulary mean scores in the productive recall test. Groups C (M=10.03) and A (M=3.35) had the highest and lowest means respectively on the delayed post-tests in-sum. However, the findings in Table 4.1 were merely based on the raw scores (i.e., 33 test items), and were only descriptively interpreted. In order to statistically measure the groups' changes across test sessions from the pre-test in-sum to the immediate post-test in-sum, and to make sure if the four groups were significantly different from one another on the immediate post-tests, ANCOVA was utilized. The scores of the immediate post-tests in-sum were considered as one dependent variable (DV), the four groups were regarded as one independent variable (IV), and the effect of the initial pre-test differences acted as the covariate. Before conducting ANCOVA, the assumption of homogeneity of variance was ensured via Levene's test (See Appendix R – Table R.2), and the result showed that the assumption was not observed ($p < .05$); thus, the alpha level was set at .025 (Tabachnick & Fidell, 2013)[40]. Table 4.2 presents the main ANCOVA results for the PR vocabulary test in regards to short-term word retention, comparing pre-test scores in-sum to the immediate post-test scores in-sum.

Table 4. 2 Tests of Between-participant Effects: ANCOVA (Pre-test in-sum to Immediate post-test in-sum)

Dependent variable: immediate post-test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	6022.767[a]	4	1505.69	18.39	.000	.419
Intercept	39007.575	1	39007.57	476.53	.000	.824
Group*pre-test	132.290	3	44.09	.53	.662	
Pre-test	.011	1	.01	.00	.991	.000
Group	6020.427	3	2006.80	24.51	.000	.419
Error	8349.476	102	81.85			
Total	66557.000	107				
Corrected Total	14372.243	106				

a. R Squared = .419 (Adjusted R Squared = .396)

The third row in Table 4.2 indicates that the assumption of homogeneity of regression slopes was also met ($p > .05$). The fourth row (Pre-test) indicates that the 4 groups were not significantly different from each other on the pre-test (i.e., covariate) ($p > .05$). Finally, the fifth row (Group) compares groups A (control), B (TAV), C (AVT), and D (VTA) in terms of their immediate post-tests in-sum. The findings show that the four groups' means were significantly different from each other on the immediate post-tests in-sum ($p < .05$) with a large effect size ($\eta p^2 = .41$), taking into account the covariate.

[40] For the discussion on ANCOVA assumptions, see methodology chapter.

As an effect size measure, I used partial eta squared (ηp^2), which refers to the proportion of total variance explained by an effect of the independent variable (Pallant, 2013; Richardson, 2011). I also relied on Cohen's (1988) guideline for interpreting the effect size range: small ηp^2/effect size = .01, medium ηp^2/effect size = .06, and large ηp^2/effect size = .14 (Pallant, 2013).

In order to see where exactly the significant differences lay, a post hoc pairwise comparison was conducted among the groups, using the Sidak adjustment test for multiple comparisons[41] (See Table 4.3).

Table 4. 3 Pairwise Comparison (Pre-test in-sum to Immediate post-test in-sum: PR* Test)

Dependent variable: immediate. post-test

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
Gr. A (Control)	Gr. B (TAV)	-13.49*	2.62	.000	-20.529	-6.453
	Gr. C (AVT)	-19.33*	2.60	.000	-26.322	-12.347
	Gr. D (VTA)	-19.58*	2.54	.000	-26.423	-12.751
Gr. B (TAV)	Gr. A (Control)	13.49*	2.62	.000	6.453	20.529
	Gr. C (AVT)	-5.84	2.46	.111	-12.446	.758
	Gr. D (VTA)	-6.09	2.47	.089	-12.739	.547
Gr. C (AVT)	Gr. A (Control)	19.33*	2.60	.000	12.347	26.322
	Gr. B (TAV)	5.84	2.46	.111	-.758	12.446
	Gr. D (VTA)	-.25	2.42	1.000	-6.771	6.266
Gr. D (VTA)	Gr. A (Control)	19.58*	2.54	.000	12.751	26.423
	Gr. B (TAV)	6.09	2.47	.089	-.547	12.739
	Gr. C (AVT)	.25	2.42	1.000	-6.266	6.771

*PR stands for productive recall test.
Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

As shown in Table 4.3, groups B (TAV), C (AVT), and D (VTA) were significantly different from group A (control) in their immediate post-test score with

[41] Sidak adjustment was used to avoid type II error (Field, 2005) for all the pairwise comparisons in this study.

larger mean differences ($p < .05$), meaning the experimental groups' performance has significantly increased on the immediate post-test in comparison to their performance on the pre-test ($p <.05$); and, they have also performed significantly better than group A (control) on the same test. The finding also signifies that the glossing technique was significantly more effective than the non-glossing strategy for L2 learners' short-term word retention. In addition, the experimental groups (B, C, & D) were not significantly different from each other on the immediate post-test in-sum ($p > .05$), meaning the glossing technique was similarly effective for L2 learners' short-term word learning and recollection. Table 4.4 also presents the adjusted means of the four groups on the immediate post-test in-sum after taking into account the effect of the covariate.

Table 4. 4 Adjusted Means after Controlling the Covariate Effect (Pre-test to immediate test in-sum): PR* Test

Dependent Variable: Immediate. Post-test

Group	Mean*	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Gr. A (Control)	8.18[a]	1.93	4.35	12.01
Gr. B (TAV)	21.67[a]	1.75	18.20	25.14
Gr. C (AVT)	27.51[a]	1.74	24.06	30.97
Gr. D (VTA)	27.77[a]	1.68	24.43	31.10

a. Covariates appearing in the model are evaluated at the following values: Pretest = .6449.
*Mean refers to each group's performance on the immediate post-test in-sum. *PR stands for productive recall test.

According to Table 4.4, group A (control) had the lowest mean score (M=8.18) in the immediate post-test in comparison to the other three experimental groups; and the mean scores of groups C (AVT) and D (VTA) were very close to each other; whereas group B (TAV) had the lowest mean (M=21.67) compared to groups C and D.

Overall, preliminary checks were conducted to ensure that there was no violation of the assumptions of ANCOVA. After adjusting for the initial pre-test differences, and according to the findings in Table 4.2, there was a statistically significant difference among the four groups in the immediate post-test scores in sum, $F_3 = 24.51$, $p = .000$, $\eta p^2 = .41$, meaning the vocabulary technique of glossing was significantly more effective than the non-glossing vocabulary strategy for participants' short-term word retention in the experimental groups when compared to the control group. Further, there was no significant difference among the means of the three experimental groups ($p > .05$), which means groups B (TAV), C (AVT), and D (VTA) performed similarly on their immediate post-test in-sum.

Immediate Post-test in-sum to Delayed post-test in-sum (Within-participant Comparison)

***1 (2)** How does glossing affect L2 learners' vocabulary learning in short-term versus long-term?*

The purpose of this comparison was to investigate if the target glossed words were retained from short-term to long-term (i.e., immediate to delayed post-tests in-sum); and if yes, which gloss order was more effective for this short versus long-term word retention. A within-participant paired samples t-test was conducted, and each group's mean score was compared from the immediate post-test in-sum to the delayed post-test in-sum. As previously shown in Table 1, group A (control) was the only group who attained the lowest mean score on the delayed post-test in-sum (M=3.35) in comparison to the three experimental groups. Also, groups B (TAV), C (AVT), and D (VTA) decreased in their performances from the immediate post-test in-sum to the delayed post-test in-sum. The declines were, however, statistically measured via paired samples t-test. Table 4.5 shows the findings of the descriptive statistics of the paired samples (the means of groups are out of 33); and Table 4.6 presents the main findings of the paired samples t-test.

Table 4. 5 Paired Samples Descriptive Statistics (Immediate Post-test in-sum to Delayed post-test in-sum):PR* Test

Group			Mean	N	Std. deviation	Std. Error Mean
Gr. A (Control)	Pair 1	Immediate post-test	8.40	20	3.70	.82
		Delayed post-test	3.35	20	3.15	.70
Gr. B (TAV)	Pair 1	Immediate post-test	22.14	27	11.16	2.14
		Delayed post-test	6.70	27	4.73	.91
Gr. C (AVT)	Pair 1	Immediate post-test	27.24	25	8.88	1.77
		Delayed post-test	10.76	25	8.20	1.64
Gr. D (VTA)	Pair 1	Immediate post-test	28.62	27	9.67	1.86
		Delayed post-test	7.59	27	6.17	1.18

*PR stands for productive recall test.

Table 4. 6 Paired Samples t-test (Immediate post-test in-sum to Delayed post-test in-sum, mean out of 33):PR* Test

Group			Paired Differences					t	df	Sig. (2-tailed)
			Mean	Std. deviation	Std. error mean	95% Confidence Interval of the Difference				
						Lower	Upper			
Group A (Control)	Pair 1	Immediate post-test- Delayed post-test	5.05	4.95	1.10	2.72	7.37	4.55	19	.000
Group B (TAV)	Pair 1	Immediate post-test- Delayed post-test	15.44	12.76	2.45	10.39	20.49	6.28	26	.000
Group C (AVT)	Pair 1	Immediate post-test- Delayed post-test	16.48	12.38	2.47	11.36	21.59	6.65	24	.000
Group D (VTA)	Pair 1	Immediate post-test- Delayed post-test	21.03	11.97	2.30	16.29	25.77	9.12	26	.000

*PR stands for productive recall test.

According to Table 4.5, the mean scores of groups A (control), B (TAV), C (AVT), and D (VTA) on the delayed post-test in-sum were lower than those of the immediate post-test in-sum. The findings in Table 4.6 also indicate that there was a significant difference between the scores of each group from the immediate post-test in-sum to the delayed post-test in-sum ($p < .05$), meaning the participants in the experimental groups (i.e., B, C, & D) did not do well in the delayed vocabulary test, and their performance atrophied in comparison to the immediate post-test.

A detailed look at the paired samples t-test table (Table 4.6) revealed that there was a statistically significant decrease in the vocabulary scores of group A participants (control) from the immediate post-test in-sum (M=8.40, SD=3.70) to the delayed post-test in-sum (M=3.35, SD= 3.15), t (19) =4.55, p < .000 (two-tailed). The mean decrease in the paired comparison was 5.05 with a 95% confidence interval ranging from 2.72 to 7.37. In group B (TAV), too, a statistically significant decrease was observed in their vocabulary scores from immediate (M=22.14, SD=11.16) to delayed post-test (M=6.70, SD=4.73), t (26) = 6.28, p < .000 (two-tailed), and the mean decrease in the paired comparison was 15.44 with a 95% confidence interval ranging from 10.39 to 20.49. Group C (AVT) showed the same statistically significant declining trend in their vocabulary scores from immediate (M=27.24, SD=8.88) to the delayed post-test (M=10.76, SD=8.20), t (24) = 6.65, p < .000 (two-tailed), and the mean decrease in the paired comparison was 16.48 with a 95% interval ranging from 11.36 to 21.59. Likewise, group D (VTA) revealed a statistically significant drop in their vocabulary scores from the immediate (M=28.62, SD=9.67) to delayed post-test (M=7.59, SD=6.17), t (26) = 9.12, p < .000 (two-tailed) with the mean decrease of 21.03 in the paired comparison and a 95% confidence interval ranging from 16.29 to 25.77.

Thus, it can be concluded that the participants in three experimental groups (B, C, & D) failed to retain a significant number of the target glossed words in long-term as compared to the short-term, and a significant attrition of the learned words in the delayed vocabulary test is revealed; meaning the glossing strategy did not significantly contribute to the word retention of the L2 learners in long-term in comparison to their short-term. Yet, among the experimental groups, the participants in group D (i.e., VTA) showed more attrition of the learned words in delayed test, t (26) = 9.12, $p < .05$, $d = .7$ (large effect size)[42] than the other groups; whereas, group B (i.e., TAV) revealed less attrition of the learned words in the same test, t (26) = 6.28, $p < .05$, $d = .5$ (large effect size) (See Table 4.6).

Pre-test in-sum to Delayed post-test in-sum (Within-participant Comparison)

1 (3) *Does glossing have any significant impact on L2 learners' vocabulary learning for long-term word retention?*

The purpose of this comparison was to investigate L2 learners' vocabulary learning and word retention in the long-term. In this analysis, the pre-test score in-sum was compared to the delayed post-test score in-sum without including the immediate post-test score in-sum. Each group was compared separately from each other (i.e., within-participant comparison) across the two test sessions (pre-and delayed). Table 4.7 presents the descriptive statistics, showing the mean score of each group on the delayed post-test in-sum. The means are out of 33.

[42] See Pallant (2013, p. 256) for the calculation of eta squared (d) for paired samples t-test.

Table 4. 7 Descriptive Statistics (Comparing Pre-test in-sum to Delayed post-test in-sum): PR Test

Dependent variable: delayed post-test. Productive Recall (PR) test

Group	Mean (pre-test)	Mean (delayed)	Std. deviation	N
Gr. A (Control)	.75	3.35	3.15	20
Gr. B (TAV)	.12	6.92	4.79	28
Gr. C (AVT)	.63	10.03	8.30	27
Gr. D (VTA)	1.03	7.32	6.22	28

Group A (control) achieved the lowest mean (M=3.35) in comparison to groups B (TAV), C (AVT), and D (VTA) on the delayed post-test in-sum. Also, group C had the highest mean (M=10.03); and group B had the lowest (M=6.92) among the three experimental groups. However, the mean differences, in the descriptive table, were statistically measured to ascertain if the four groups were significantly different from one another on the delayed post-test in-sum. ANCOVA was utilized for this purpose with the effect of the initial pre-test scores in-sum as a covariate. The score on the delayed post-test in-sum formed one DV, and the groups were considered as one IV. The normality was already ensured (See Table 4.1), and the homogeneity of variances was checked via Levene's test (See Appendix R – Table R.3). The result showed that Levene's assumption was not observed for this analysis ($p < .05$); thus, as before, the alpha level was set at .025 (Tabachnick & Fidell, 2013). Table 4.8 demonstrates the main ANCOVA results in regard to the long-term word retention, where the pre-test score in-sum was compared to the delayed post-test score in-sum without including the immediate post-test score in-sum.

Table 4. 8 Tests of Between-participant Effects: ANCOVA (Pre-test in-sum to Delayed post-test in-sum, PR Test)

Dependent variable: delayed post-test. Productive Recall (PR) test

Source	df	F	Sig.	Partial eta squared
Corrected Model	4	7.77	.000	.24
Intercept	1	84.79	.000	.46
Group*pre-test	3	3.25	.025	
Pre-test	1	15.09	.000	.13
Group	3	6.36	.001	.16
Error	98			
Total	103			
Corrected Total	102			

a. R Squared = .241 (Adjusted R Squared = .210)

Like all ANCOVA tables before, the assumption of homogeneity of regression slopes was observed ($p > .05$) (third row in Table 4.8). The fourth row (Pre-test) also shows that the 4 groups were significantly different from each other in the pre-test (i.e., covariate) ($p < .05$); and the fifth row (Group) compares the control and three experimental groups in terms of their delayed post-tests. The result showed that groups A (control), B (TAV), C (AVT), and D (VTA) were significantly different from each other on the delayed post-tests in-sum ($p < .05$) with a large effect size ($\eta p^2 = .16$), taking into account the initial pre-test differences (i.e., covariate). However, in order to exactly locate the area of differences among the four groups, a post hoc pairwise comparison was conducted, and the four groups were compared together in terms of their sum of the delayed post-test score (See Table 4.9).

Table 4. 9 Pairwise Comparison (Pre-test in-sum to Delayed post-test in-sum): PR Test

Dependent variable: delayed post-test. Productive Recall (PR) test

(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-4.93*	.027
	Gr. C (AVT)	-7.27*	.000
	Gr. D (VTA)	-3.80	.138
Gr. B (TAV)	Gr. A (Control)	4.93*	.027
	Gr. C (AVT)	-2.34	.574
	Gr. D (VTA)	1.13	.978
Gr. C (AVT)	Gr. A (Control)	7.27*	.000
	Gr. B (TAV)	2.34	.574
	Gr. D (VTA)	3.47	.150
Gr. D (VTA)	Gr. A (Control)	3.80	.138
	Gr. B (TAV)	-1.13	.978
	Gr. C (AVT)	-3.47	.150

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

The findings show that there was no significant difference between groups B (TAV), C (AVT), and D (VTA) on the delayed post-test ($p > .05$). Also, groups B and C performed significantly better on the delayed post-test only in comparison to group A (control) ($p < .05$), and not group D ($p > .05$). In addition, there was no significant difference between groups A and D ($p > .05$) in the scores of the delayed post-test. In other words, group D did not show a significant long-term word retention. Thus, it can be concluded that the three experimental groups were not significantly different on the delayed post-test ($p > .05$), meaning groups B (TAV), C (AVT), and D (VTA) performed similarly on their delayed post-test.

Table 4.10 also presents the adjusted means of the four groups on the delayed post-test score in-sum after taking into account the effect of the covariate. From Table 4.10, it can be found that group A (control) received the lowest mean score (M= 2.87) on the delayed post-test in comparison to the three experimental groups; and the performance of group C (AVT) was better (M=10.14) than groups B (TAV) and D (VTA).

Table 4. 10 Adjusted Means after Controlling the Covariate (Pre-test in-sum to Delayed post-test in-sum)

Dependent variable: delayed post-test. Productive Recall (PR) test

Group	Mean	Std. error
Gr. A (Control)	2.87[a]	1.27
Gr. B (TAV)	7.80[a]	1.09
Gr. C (AVT)	10.14[a]	1.09
Gr. D. (VTA)	6.67[a]	1.08

a. Covariates appearing in the model are evaluated at the following values: Pretest = .6214.

Altogether, after preliminary checks on the assumptions of ANCOVA and adjusting the means for initial pre-test differences, according to Table 4.8, there was a statistically significant difference among the four groups in the delayed post-test scores in-sum, $F_3 = 6.37$, $p = .001$, $\eta p^2 = .16$ (large effect size); however, it can be concluded, with caution, that the glossing technique of vocabulary instruction was partially effective for L2 learners' long-term word retention, because groups B and C showed significant differences on their delayed post-test scores in comparison to groups A and D. Besides, the groups can be ranked from the lowest performance to the highest performance on the delayed post-test as groups A (control), D (VTA), B (TAV), and C (AVT).

It is noteworthy to mention that the analyses above were pertinent to the in-sum comparisons of glossing on L2 learners' vocabulary learning in terms of short and long-term word retention. Yet, in all these comparisons, the four groups were compared separately from one another; and the sub pre/immediate and delayed post-tests were not included. In other words, the comparisons did not exactly specify which glossing mode(s) helped L2 learners to learn the target glossed words significantly better than another in terms of both short and long-term word retention. To address this issue, detailed between-participant comparisons, including sub-test comparisons, were performed. The next section sheds light on the findings.

Sub-test Comparisons

As stated earlier, in order to do the sub-test comparisons, the sub pre/immediate and delayed post-tests were created. The sub-test included the sub component of each pre/delayed test. The purpose of creating the sub-test was to bring more clarity to the findings. The sub-test comparisons addressed the following research question:

2. *Do different glossing modes (L2 definition alone, L2 definition and audio glossing or L2 definition and video/animation glossing) have any significant impact on L2 learners' vocabulary learning and short and long-term word retention?*

This question includes two sub research questions, which are analyzed and responded subsequently. Additionally, another analysis was done to compare participates' scores from sub immediate to sub delayed post-tests for short versus long-term effect (i.e., RQ. 2.2). For the purpose of clarity and general overview of the sub-test comparisons, Table 4.11 presents the descriptive statistics of the four groups (control & three experimental groups), showing their performance from the sub pre-test to the sub delayed post-test across test sessions. The means represent each group's performance on the sub-tests.

Table 4. 11 Descriptive Statistics of Groups from Sub pre-tests to Sub delayed post-tests (PR test)

Group		N	Mean	Std. deviation	Skewness		Kurtosis	
						Std. error		Std. error
Gr. A (Control)	Pre.Sub1	24	.33	.70	1.88	.472	2.09	.918
	Pre.Sub2	24	.37	.76	1.70	.472	1.13	.918
	Pre.Sub3	24	.04	.20	4.89	.472	24.00	.918
	Im. Post.Sub1	24	1.33	1.30	.58	.472	-.60	.918
	Im. Post.Sub2	24	2.33	2.46	.47	.472	-1.47	.918
	Im. Post.Sub3	22	4.31	1.72	.36	.491	-.64	.953
	Del.Post.Sub1	20	1.05	1.63	1.82	.512	3.44	.992
	Del.Post.Sub2	20	.90	1.02	.21	.512	-2.18	.992
	Del.Post.Sub3	20	1.40	1.46	.55	.512	-.83	.992
Gr. B (TAV)	Pre.Sub1	39	.02	.16	6.24	.378	39.00	.741
	Pre.Sub2	39	.00	.00	.	.	.	.
	Pre.Sub3	39	.10	.44	4.23	.378	16.77	.741
	Im.Post.Sub1	30	6.13	5.41	1.28	.427	1.817	.833
	Im.Post.Sub2	30	6.06	5.44	1.05	.427	.273	.833
	Im.Post.Sub3	31	8.77	4.60	.68	.421	-.392	.821
	Del.Post.Sub1	28	1.84	2.44	1.37	.378	1.734	.741
	Del.Post.Sub2	28	1.15	1.76	1.74	.383	2.997	.750
	Del.Post.Sub3	28	2.78	1.68	-.03	.441	-.722	.858
Gr. C (AVT)	Pre.Sub1	36	.22	.92	4.05	.393	15.26	.768
	Pre.sub2	36	.41	.73	1.46	.393	.60	.768
	Pre.sub3	36	.00	.00	.	.	.	.
	Im. Post.Sub1	33	9.45	5.60	-.48	.409	-1.07	.798
	Im. Post.Sub2	30	9.43	4.24	.83	.427	.53	.833
	Im. Post.Sub3	30	9.13	5.51	.17	.427	-1.09	.833
	Del.Post.Sub1	27	2.25	3.39	2.44	.448	7.03	.872
	Del.Post.Sub2	27	4.70	3.99	.94	.448	.96	.872

	Del.Post.Sub3	27	3.07	2.40	.59	.448	.31	.872
Gr. D (VTA)	Pre.Sub1	33	.48	1.12	2.30	.409	4.51	.798
	Pre.Sub2	33	.42	.83	1.47	.409	.18	.798
	Pre.Sub3	33	.12	.48	3.86	.409	13.73	.798
	Im. Post.Sub1	32	11.78	5.94	-.45	.414	-.86	.809
	Im. Post.Sub2	31	7.19	5.04	.56	.421	-.44	.821
	Im. Post.Sub3	31	9.19	4.88	.43	.421	-.80	.821
	Del.Post.Sub1	28	3.21	2.89	.83	.441	.20	.858
	Del.Post.Sub2	28	2.32	2.98	1.51	.441	1.82	.858
	Del.Post.Sub3	28	1.78	2.16	1.07	.441	.12	.858

Note: Im. Post. Sub stands for sub immediate post-test and Del. Post. Sub stands for sub delayed post-test. Number of test items: sub-pre/immediate & post-test 1 = 12 items; sub-pre/immediate & delayed 2 = 11 items; &
sub pre/immediate & delayed 3 = 10 items.

According to Table 4.11, group A (control) achieved the lowest mean scores on both the sub immediate and sub delayed post-tests, compared to groups B (TAV), C (AVT), and D (VTA). The three experimental groups (B, C, & D) also performed better than group A in all the three sub immediate post-tests; whereas their performance declined on the corresponding sub delayed post-tests. The next section presents the inferential findings regarding all four group changes across sessions (i.e., between-participant comparisons), and examines if the mean differences were statistically significant among the groups from sub pre-tests to sub delayed post-tests. This between-participant comparison also investigates which glossing mode (s) was significantly effective for L2 learners' short and long-term word learning and retention. All the relevant assumptions for sub-test analyses were also checked and ensured.

Sub pre-test to Sub immediate post-test (Between-participant Comparison)

2 (1) *Which glossing mode (s) contributes significantly to L2 learners' short-term word learning and retention?*

To do the analysis, all four groups (control & three experimental groups) were compared together from the sub pre-tests to the corresponding sub immediate post-tests, using ANCOVA with the effect of the sub pre-test scores as the covariate; sub immediate post-test scores as one DV and groups as one IV. Checking the assumption of homogeneity of variances (See Appendix R – Table R.4), it was found that the assumption was not observed for this analysis (p < .05); thus, the alpha level was set at .025 (Tabachnick & Fidell, 2013).

Table 4. 12 Tests of Between-participant Effects: ANCOVA (Sub pre-tests to Sub immediate post-tests. Productive Recall (PR) Test

	Comparing pre1-im1				Comparing pre2-im2				Comparing pre3-im3			
Source	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared
Corrected Model	4	16.71	.000	.37	4	8.36	.000	.23	4	5.29	.001	.16
Intercept	1	223.02	.000	.66	1	177.96	.000	.61	1	313.55	.000	.74
Group*pre-sub1	3	.41	.742		2	1.24	.293		2	1.28	.282	
Pre-sub1	1	1.86	.175	.01	1	.00	.938	.000	1	1.96	.164	.01
Group	3	22.12	.000	.36	3	11.14	.000	.23	3	6.31	.001	.14
Error	114				110				109			
Total	119				115				114			
Corrected Total	118				114				113			

Table 4.12 presents the main findings of three ANCOVAs comparing each sub pre-test to the relevant sub immediate post-test across the four groups (A, B, C, & D). The assumption of homogeneity of regression slopes in the third row was also met (p > .05). The fourth row (Pre-sub1) in Table 4.12 shows that the four groups were not

significantly different from each other in the pre-test (i.e., covariate) (p > .05). Finally, the fifth row (Group) compares the groups (control & three experimental groups) in terms of their sub immediate post-tests. The results showed that groups A (control), B (TAV), C (AVT), and D (VTA) were significantly different from each other on the sub immediate post-tests (p < .05) with large effect sizes (ranging from .14 to .36). However, to examine where exactly the significant differences existed, a post hoc pairwise comparison was conducted (See Table 4.13).

Table 4. 13 Pairwise Comparison (Sub pre-test to Sub immediate post-test: Between-participant Effects): PR Test

Dependent variable: impost-tests 1, 2, & 3. Productive Recall (PR) test

		Comparing pre1-im1		Comparing pre2-im2		Comparing pre3-im3	
(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-4.57*	.009	-3.71*	.025	-4.42*	.004
	Gr. C (AVT)	-8.05*	.000	-7.10*	.000	-4.89*	.001
	Gr. D (VTA)	-10.57*	.000	-4.86*	.001	-4.72*	.002
Gr. B (TAV)	Gr. A (Control)	4.57*	.009	3.71*	.025	4.42*	.004
	Gr. C (AVT)	-3.48*	.047	-3.38*	.034	-.47	.999
	Gr. D (VTA)	-6.00*	.000	-1.15	.919	-.30	1.000
Gr. C (AVT)	Gr. A (Control)	8.05*	.000	7.10*	.000	4.89*	.001
	Gr. B (TAV)	3.48*	.047	3.38*	.034	.47	.999
	Gr. D (VTA)	-2.52	.261	2.23	.303	.17	1.000
Gr. D (VTA)	Gr. A (Control)	10.57*	.000	4.86*	.001	4.72*	.002
	Gr. B (TAV)	6.00*	.000	1.15	.919	.30	1.000
	Gr. C (AVT)	2.52	.261	-2.23	.303	-.17	1.000

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

The pairwise comparison reveals that group A (control) had significantly lower mean differences in all the sub immediate post-tests (p > .05), and the three experimental groups achieved significantly higher means than the control group (p < .05); meaning groups B, C, and D outperformed group A (control) in their performance on all the immediate post-tests. Consequently, different glossing modes were significantly more effective than the non-glossing mode for L2 learners' short-term vocabulary learning in this between-participant comparison, because according to Table 4.13, there were some significant differences among the performance of groups B, C, and D in the sub immediate post-tests ($p < .05$).

In the sub-test 1 comparison, groups C (AVT) and D (VTA) performed significantly better than group B (TAV) ($p < .05$), and achieved higher mean differences; whereas group B was only significantly better than group A ($p < .05$). Group B received the glossed words via the single mode of L2 definition alone. Furthermore, the same sub-test comparison shows that there was no significant difference between groups C and D on the sub immediate post-test 1 ($p > .05$); meaning the participants in these two groups performed similarly on immediate vocabulary tests. Groups C and D received the glossed words via the dual modes of L2 definition and audio glossing and L2 definition and video/animation glossing, respectively, on the first day of instruction. Therefore, based on the sub-test 1 comparison in Table 4.13, the dual glossing modes (i.e., L2 definition and audio glossing as well as L2 definition and video/animation glossing) were significantly more effective than the single glossing mode or no glossing mode for L2 learners' short-term word learning.

In sub-test 2 comparison, groups B (TAV) and D (VTA) were not significantly different from each other ($p > .05$); Likewise, groups C (AVT) and D (VTA) did not perform significantly different in their sub immediate post-test 2 ($p > .05$); however, group C (AVT) acted significantly better than groups B and D. This group was instructed via the bimodal glossing of L2 definition and video/animation on day 2. Therefore, based on sub-test 2 comparison in Table 4.13, the dual mode of L2 definition and video/animation glossing was significantly more effective than the two other glossing modes (i.e., L2 definition and audio glossing OR L2 definition alone), or no glossing mode for L2 learners' word learning.

Finally, in the third sub-test comparison, surprisingly, there was no significant difference among the performance of groups of B (TAV), C (AVT), and D (VTA) (p > .05).

Table 4.14 presents the adjusted means of the 4 groups (control & three experimental) on sub immediate post-tests while controlling the covariate. As shown, group A (control) had the lowest mean scores on all sub immediate post-tests in comparison to groups B (TAV), C (AVT) and D (VTA). Besides, group D (VTA) in sub-test 1 comparison, and group C (AVT) in sub-tests 2 and 3 comparisons achieved higher mean scores on sub immediate post-tests.

Table 4. 14 Adjusted Means after Controlling the Covariate Effect (Sub pre-test to Sub immediate test) (PR Test)

Dependent variable: immediate post-tests 1, 2, & 3. Productive Recall (PR) test

	Comparing pre1-im1		Comparing pre2-im2		Comparing pre3-im3	
Group	Mean	Std. error	Mean	Std. error	Mean	Std. error
Gr. A (Control)	1.37[a]	1.03	2.33[a]	.93	4.34[a]	.97
Gr. B (TAV)	5.94a	.93	6.05[a]	.85	8.76[a]	.81
Gr. C (AVT)	9.42[a]	.88	9.43[a]	.83	9.24[a]	.83
Gr. D (VTA)	11.95[a]	.90	7.20[a]	.82	9.07[a]	.82

According to the between-participant analysis (Table 4.12), findings show that there were statistically significant differences from sub pre-test 1 to sub immediate post-test 1, $F_3 = 22.12$, $p = .000$, $\eta p^2 = .36$ (large effect size); from sub pre-test 2 to sub immediate post-test 2, $F_3 = 11.14$, $p = .000$, $\eta p^2 = .23$ (large effect size); and finally, from sub pre-test 3 to sub immediate post-test 3, $F_3 = 6.31$, $p = .001$, $\eta p^2 = .14$ (large effect size). Therefore, different glossing modes were significantly effective for L2 learners' word learning and retention in short-term. The bimodal glossing of L2 definition and audio glossing, and L2 definition and video/animation glossing were significantly more effective than single glossing mode or no glossing for short-term retention of the participants on day 1. Only the bimodal glossing of L2 definition and video/animation was significantly effective for learners' short-term retention on day 2; and all three glossing modes were not significantly different from one another for participants' word learning in short-term on day 3.

Sub immediate post-tests to Sub delayed post-tests (Between-participant Comparison)

2 (2) *Which glossing mode (s) affect L2 learners' vocabulary learning in short-term versus long-term?*

In order to examine which glossing mode(s) led to less attrition of the target glossed words in comparison to the learners' short-term word retention, the four groups were compared from sub immediate post-tests to the corresponding sub delayed post-tests. Figure 4.2 displays the visual line graph of the four groups in each sub-test from immediate to delayed post-test.

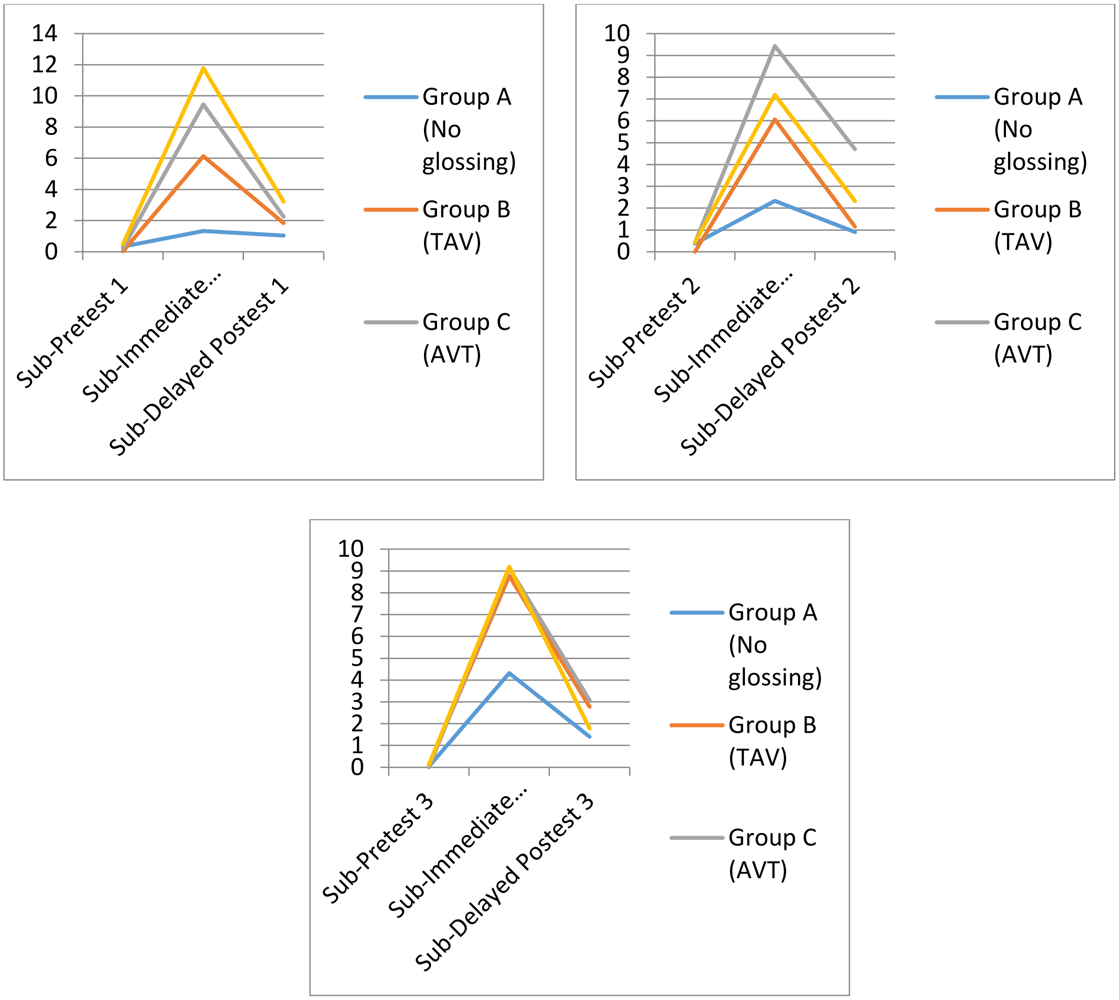


Figure 4. 2 Trend of group changes across test sessions (PR Test)

As visually displayed, the groups declined from sub immediate post-tests to sub delayed post-tests. Group A (control) performed the lowest in comparison to groups B (TAV), C (AVT), and D (VTA) on all sub delayed tests. The performance of the participants in groups B, C, and D also decreased on the delayed vocabulary tests in comparison to their immediate post-tests. The inferential test of ANCOVA was conducted to compare the scores from the sub immediate post-test to the corresponding sub delayed post-tests across the groups while taking into account the effect of the covariate (i.e., sub immediate post-test differences). Preliminary checks were done on

ANCOVA assumptions, and the result of homogeneity of variances showed violation for sub comparison 3 (sub immediate 3 to sub delayed 3) ($p < .05$) (See Appendix R – Table R.5); thus, the alpha level was set at .025 (Tabachnick & Fidell, 2013) for this analysis. Also, homogeneity of regression slopes was checked and assumed ($p > .05$) in the third row of Table 4.15, which presents the main results of the three ANCOVAs for this comparison.

Table 4. 15 Tests of Between-participant Effects: ANCOVA (Sub immediate to Sub delayed tests across Groups. Productive Recall (PR) test

	Comparing Im1-del1				Comparing Im2-del2				Comparing Im3-del3			
Source	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared
Corrected Model	4	2.19	.075	.08	4	2.82	.029	.10	4	8.00	.000	.24
Intercept	1	33.94	.000	.25	1	26.04	.000	.21	1	24.14	.000	.19
Group* Im.Post1	3	.55	.646		3	1.24	.299		3	.86	.460	
Im.Post1	1	1.49	.225	.01	1	.25	.614	.00	1	.09	.765	.00
Group	3	2.83	.042	.07	3	3.18	.027	.09	3	8.60	.000	.21
Error	100				97				97			
Total	105				102				102			
Corrected Total	104				101				101			

The fourth row (Im. Post1) indicates that the 4 groups (control & three experimental groups) were not significantly different from each other on the sub immediate post-tests (i.e., covariate) ($p > .05$); and the fifth row (Group) compares groups A (control), B (TAV), C (AVT), and D (VTA) in terms of their sub delayed vocabulary tests. As this between-participant comparison demonstrates, groups A, B, C, and D were significantly different from each other in their performance on all sub delayed post-tests ($p < .05$) with medium to large effect sizes (ranging from .07 to .21). However, to locate the significant differences among the four groups, post hoc pairwise comparisons were conducted, and the control and three experimental groups were compared with one another in terms of their sub delayed post-tests (See Table 4.16). This table also identifies which glossing mode(s) resulted in less attrition of the learned target words from short-term to long-term.

Table 4. 16 Pairwise Comparison (Sub immediate post-tests to Sub delayed post-tests across Groups) (PR Test)

Dependent variable: immediate post-tests 1, 2, & 3. Productive Recall (PR) test

		Comparing Im1-del1		Comparing Im2-del2		Comparing Im3-del3	
(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-1.65	.259	-1.19	.306	-.43	.996
	Gr. C (AVT)	-1.67	.329	-1.56	.081	-3.97*	.000
	Gr. D (VTA)	-2.86*	.027	-.27	.998	-1.42	.475
Gr. B (TAV)	Gr. A (control)	1.65	.259	1.19	.306	.43	.996
	Gr. C (AVT)	-.01	1.000	-.37	.984	-3.54*	.000
	Gr. D (VTA)	-1.21	.563	.91	.438	-.98	.704
Gr. C (AVT)	Gr. A (control)	1.67	.329	1.56	.081	3.97*	.000
	Gr. B (TAV)	.01	1.000	.37	.984	3.54*	.000
	Gr. D (VTA)	-1.19	.541	1.29	.107	2.55*	.008
Gr. D (VTA)	Gr. A (control)	2.86*	.027	.27	.998	1.42	.475
	Gr. B (TAV)	1.21	.563	-.91	.438	.98	.704
	Gr. C (AVT)	1.19	.541	-1.29	.107	-2.55*	.008

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

In comparison between the sub immediate and delayed post-test 1, the performance of the participants in group A (control) was not significantly different from those in groups B (TAV) and C (AVT) ($p > .05$); but significantly different from group D (VTA) ($p < .05$). In other words, L2 learners in group D did significantly better than the learners in group A on sub delayed post-test 1, and retained the learned words better from short-term (sub immediate post-test 1) to long-term (sub delayed post-test 1) (i.e., less attrition of the words in long-term in comparison to their short-term), $F3 = 2.83$, $p = .04$, $\eta p^2=.07$ (medium effect size). The participants in group D (VTA) received the glossed words via L2 definition and video/animation glossing on day 1. Hence, this glossing mode helped them to retain the words better from short-term to long-term than the glossing modes of L2 definition and audio glossing or L2 definition alone. Also, the three experimental groups were not significantly different from each other ($p > .05$). Therefore, the findings should be interpreted with caution here.

In the second comparison between sub immediate and the delayed post-test 2, the four groups were not significantly different from each other, and failed to retain the words in long-term in comparison to the short-term to a significant amount ($p > .05$), $F3 = 3.18$, $p = .02$, $\eta p^2 = .09$ (medium effect size). Thus, different glossing modes were not significantly effective for L2 learners' long-term word retention, compared to the short-term word retention on day 2.

Finally, according to the third sub-test comparison, the participants in group C (AVT) were the only ones who showed a significant difference on the sub delayed vocabulary test 3, in comparison to groups A (control), B (TAV) and D (VTA) ($p < .05$), meaning they could retain the words significantly better from short-term to long-term; and thus, showed less attrition of the learned words, $F3 = 8.60$, $p = .000$, $\eta p^2 = .21$ (large effect size). Surprisingly, the participants in this group were instructed via the single glossing mode of L2 definition alone on day 3.

Table 4.17 presents the adjusted means of the four groups on sub delayed post-tests after considering the effect of the covariate. As shown, group A (control) received the lowest mean on all sub delayed post-tests in comparison to the three experimental groups (B, C, & D); and the mean scores of groups B, C, and D were also low when compared to their immediate post-test scores.

Table 4. 17 Adjusted Means after Controlling the Covariate Effect (Sub immediate to Sub delayed post-tests)

Dependent variable: immediate post-tests 1, 2 & 3. Productive Recall (PR) test						
	Comparing Im1-del1		Comparing Im2-del2		Comparing Im3-del3	
Group	Mean	Std. error	Mean	Std. error	Mean	Std. error
Gr. A (control)	.65[a]	.69	1.48[a]	.47	.97[a]	.66
Gr. B (TAV)	2.31[a]	.50	2.68[a]	.38	1.40[a]	.50
Gr. C (AVT)	2.32[a]	.53	3.05[a]	.38	4.94[a]	.57
Gr. D (VTA)	3.52[a]	.57	1.76[a]	.38	2.39[a]	.53

In short, the between-participant comparison (Table 4.15) shows that the four groups declined significantly from sub immediate to sub delayed post-tests ($p <.05$). Yet, in order to examine which glossing mode (s) helped L2 learners to experience less attrition of the learned words from short-term to long-term (sub immediate to sub delayed test), Table 4.16 shows that on day 1, the dual mode L2 definition and video/animation glossing contributed to L2 learners' less attrition of the glossed words; three glossing modes contributed similarly to L2 learners' attrition of the words from short-term to long-term on day 2; and interestingly, single mode of L2 definition alone was more effective than the dual glossing modes for L2 learners' less word attrition from short to long-term on day 3. The next section discusses the results in regard to long-term word learning and retention across groups.

Sub pre-test to Sub delayed post-test (Between-participant Comparison)

2 (3) *Which glossing mode (s) contributes significantly to L2 learners' long-term word learning and retention?*

The purpose of this comparison was to investigate which glossing mode (s) was significantly effective for the long-term word learning and retention of L2 learners across control (A) and three experimental groups (B, C & D) (i.e., between-participant comparison). Table 4.18 presents the descriptive statistics of this comparison, including the means of the four groups on sub-delayed vocabulary tests.

Table 4. 18 Descriptive Statistics (Sub pre-test to Sub delayed post-test Without sub-immediate) (PR Test)

Dependent variable: sub delayed post-test. Productive Recall (PR) test

	Sub-test 1			Sub-test 2			Sub-test 3		
Group	Mean	Std. deviation	N	Mean	Std. deviation	N	Mean	Std. deviation	N
Gr. A (Control)	1.05	1.63	20	.90	1.02	20	1.40	1.46	20
Gr. B (TAV)	1.84	2.44	39	1.15	1.76	38	2.78	1.68	28
Gr. C (AVT)	2.25	3.39	27	4.70	3.99	27	3.07	2.40	27
Gr. D (VTA)	3.21	2.89	28	2.32	2.98	28	1.78	2.16	28

According to Table 4.18, group A (control) received the lowest mean scores on all sub delayed post-tests, as compared to groups B (TAV), C (AVT), and D (VTA). Besides, group D on sub delayed post-test 1(M = 3.21) and group C on sub delayed post-tests 2 and 3 received the highest mean scores (M = 4.70; M = 3.07 respectively) in comparison to the other groups (A & B) on each sub delayed-test. It is interesting to note that, according to Table 4.11, participants' mean scores in all four groups (A, B, C, & D) increased from sub pre-tests to sub delayed post-tests. However, ANCOVA was conducted to examine if the mean differences among the groups were statistically significant; it compared the participants' scores on each sub pre-test to the corresponding sub delayed post-test without intervening the sub immediate post-test scores across groups. In this analysis, groups formed one IV, the sub delayed test scores formed one DV, and the initial pre-test scores acted as a covariate. All ANCOVA assumptions were also controlled; however, the result of Levene's test showed the violation of this assumption for sub-test comparison 2 ($p < .05$) (See Appendix R – Table R.6); thus, to analyze the data, the alpha level was set at .025 (Tabachnick & Fidell, 2013). The findings of the three ANCOVAs are demonstrated in Table 4.19.

Table 4. 19 Tests of Between-participant Effects: ANCOVA (Sub pre-tests to Sub delayed post-tests) (PR Test)

Dependent variable: sub delayed post-test. Productive Recall (PR) test

	Sub-test 1				Sub-test 2				Sub-test 3			
Source	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared
Corrected Model	4	9.94	.000	.26	4	8.39	.000		4	2.97	.023	
Intercept	1	44.74	.000	.29	1	57.25	.000		1	123.32	.000	
Group* Pre.Sub1	3	3.73	.013		2	.61	.542		2	.58	.559	
Pre.Sub1	1	29.46	.000	.21	1	.13	.719		1	.34	.559	
Group	3	3.17	.027	.08	3	10.74	.000	.23	3	3.91	.011	.10
Error	109				108				98			
Total	114				113				103			
Corrected Total	113				112				102			

The assumption of homogeneity of regression slopes was observed (p > .05) (third row in Table 4.19). The fourth row (Pre. Sub1) shows that the 4 groups (control & three experimental groups) were significantly different from each other on sub pre-test 1, F_3 =29.46, p = .000, ηp^2= .021; but they were not significantly different from each other on sub pre-tests 2 and 3 (p > .05). Finally, the fifth row (Group) compared the 4 groups in terms of their sub delayed post-tests. As shown, groups A (control), B (TAV), C (AVT) and D (VTA) were significantly different from each other on sub delayed post-tests (p < .05) with a small to large effect size (ranging from .1 to .08), considering the covariate. Post hoc pairwise comparisons were conducted to examine where the significant differences lay among the groups (See Table 4.20).

Table 4. 20 Pairwise Comparison (Sub pre-tests to Sub delayed post-tests) (PR Test)

Dependent variable: sub delayed post-tests. Productive Recall (PR) test

		Sub-test 1		Sub-test 2		Sub-test 3	
(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-1.46	.178	-.32	.999	-1.37	.119
	Gr. C (AVT)	-1.65	.127	-3.81*	.000	-1.69*	.030
	Gr. D (VTA)	-2.11*	.020	-1.41	.381	-.37	.988
Gr. B (TAV)	Gr. A (Control)	1.46	.178	.32	.999	1.37	.119
	Gr. C (AVT)	-.19	1.000	-3.48*	.000	-.31	.993
	Gr. D (VTA)	-.65	.871	-1.09	.544	1.00	.330
Gr. C (AVT)	Gr. A (Control)	1.65	.127	3.81*	.000	1.69*	.030
	Gr. B (TAV)	.19	1.000	3.48*	.000	.31	.993
	Gr. D (VTA)	-.45	.982	2.39*	.008	1.31	.098
Group D (VTA)	Gr. A (Control)	2.11*	.020	1.41	.381	.37	.988
	Gr. B (TAV)	.65	.871	1.09	.544	-1.00	.330
	Gr. C (AVT)	.45	.982	-2.39*	.008	-1.31	.098

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

Table 4.20 shows that, on the first sub delayed post-test, the performance of groups A (control), B (TAV) and C (AVT) were not significantly different from each other (p > .05), but group D (VTA) was significantly different from group A (p < .05); and the three experimental groups (B, C, & D) were not significantly different from each other (p > .05). In other words, only participants in group D were able to retain the target glossed words in long-term in comparison to the participants of group A ($F_3 = 3.17$, p = .02, ηp^2= .08), and not in comparison to the other groups. The participants in group D were instructed via the bimodal glossing of L2 definition and video/animation on the first day of the instruction. Thus, this glossing mode helped L2 learners to retrieve the words in long-term significantly better than the two other glossing modes of L2 definition and audio glossing or L2 definition alone, and even no glossing mode) on day 1. However, the result here should be interpreted with caution since L2 learners in group D outperformed those in group A (control) only.

On the second sub delayed post-test, the performance of the participants in groups A (control), B (TAV) and D (VTA) were not significantly different from each other ($p > .05$), but significantly different from group C (AVT), meaning group C retained the learned glossed words to a significant extent in long-term ($F_3 = 10.74$, $p = .00$, $\eta p^2 = .23$). The participants in group C received the glossed words through the bimodal glossing of L2 definition and video/animation on the second day of the instruction. Thus, like day 1, the dual glossing mode had a significantly positive impact on L2 learners' long-term word learning and retention.

Finally, on sub delayed post-test 3, there was a statistically significant difference between groups C (AVT) and A (control) ($p < .05$), meaning the participants in group C performed significantly better only in comparison to group A (control), and could retrieve the learned words in long-term ($F3 = 3.91$, $p = .01$, $\eta p^2 = .10$); and not in comparison to the two other experimental groups of B and D. Group C participants were taught the target glossed words via the single mode of L2 definition alone on the third day of the instruction. Thus, this glossing mode had a significantly positive influence on L2 learners' long-term word learning and retention. Also, there was no significant difference among groups B (TAV), C (AVT), and D (VTA) themselves ($p > .05$). Surprisingly, groups B and D did not perform significantly different from group A (control) in this test as well ($p > .05$).

Table 4.21 displays the adjusted means of the 4 groups (control & three experimental groups) on sub delayed vocabulary tests after controlling the covariate.

Group A (control) achieved the lowest mean scores on all sub delayed tests; also, group D (VTA) on sub delayed post-test 1, and group C (AVT) on sub delayed post-tests 2 and 3 received the highest mean scores (*M*= 2.84; *M*= 4.68; *M*= 3.09) respectively.

Table 4. 21 Adjusted Means after Controlling the Covariate (Sub pre-tests to Sub delayed post-tests) (PR Test)

Dependent variable: sub delayed post-tests. Productive Recall (PR) test

	Sub-test 1		Sub-test 2		Sub-test 3	
Group	Mean	Std. error	Mean	Std. error	Mean	Std. error
Gr. A (Control)	.72[a]	.542	.879[a]	.608	1.399[a]	.448
Gr. B (TAV)	2.19[a]	.391	1.201[a]	.455	2.777[a]	.379
Gr. C (AVT)	2.38[a]	.464	4.688[a]	.522	3.093[a]	.387
Gr. D (VTA)	2.84[a]	.461	2.293[a]	.517	1.777[a]	.379

Overall, the between-participant comparison (Table 4.19) shows that the control and three experimental groups were significantly different from each other on all sub delayed post-tests ($p < .05$). However, to examine which glossing mode (s) caused a significantly positive influence on L2 learners' long-term word learning and retention, the pairwise comparison (Table 4.20) revealed that the dual mode of L2 definition and video/animation glossing was effective on the first two days of the instruction (days 1 & 2); and the single mode of L2 definition alone was influential on day 3 for L2 learners' long-term vocabulary retention.

Summary of Section One

In section one, the findings regarding the productive recall vocabulary test were presented and analyzed. Two types of between and within-participant comparisons were carried out via ANCOVA and paired samples t-test respectively. The data were once analyzed with in-sum comparisons and another time with sub-test comparisons.

The results of both between and within in-sum comparisons show that: **(a)** the vocabulary technique of glossing was significantly more effective than the non-glossing strategy for participants' short-term word learning and retention; **(b)** the glossing technique did not contribute significantly to L2 learners' word retention in long-term when compared to the short-term. Yet, among the experimental groups, the participants in group D (i.e., VTA) showed more attrition of the learned words in delayed test than the other groups; whereas, group B participants (i.e., TAV) revealed less attrition of the glossed words in the same test; and finally, **(c)** the vocabulary technique of glossing was partially effective for L2 learners' long-term word learning and retention.

The results of the between sub-test comparisons revealed that: **(a)** different glossing modes were significantly effective for L2 learners' short-term word retention, and the two modes of L2 definition and audio glossing, and L2 definition and video/animation glossing were significantly more effective than the single mode of L2 definition alone or no glossing mode for learners' short-term retention on day 1; only the bimodal glossing of L2 definition and video/animation was significantly effective on day 2; and all three glossing modes were not significantly different from one another on day 3; **(b)** glossing modes were differently effective for L2 learners' attrition of the target words from short-term to long-term word retention; on day 1, dual mode of L2 definition and video/animation glossing contributed to L2 learners' less attrition of the learned words; three glossing modes contributed similarly to L2 learners' attrition of the words from short-term to long-term on day 2; and interestingly, the single mode of L2 definition alone was more effective than the two dual modes for L2 learners' less word attrition from short to long-term on day 3; and finally, **(c)** glossing modes contributed significantly to L2 learners' word learning and retention in long-term; the dual mode of L2 definition and video/animation glossing was effective on the first two days of the instruction (days 1 & 2); and the single mode of L2 definition alone was influential on day 3 for L2 learners' long-term vocabulary retention.

Finally, it can be noted that the participants in the three experimental conditions for productive recall tests significantly increased their scores from the pre-test to the immediate post-test, but also revealed a significant loss from the immediate to the delayed post-test. However, this negative effect was counterbalanced by the fact that the learners still showed a significant gain in scores from the pre-test to the delayed post-test. Thus, it can be concluded that some words learned through glossing tend to be maintained based on productive recall measures.

SECTION TWO

(Multiple-choice Productive Recognition Test)

In-Sum Comparisons

As stated earlier, the in-sum comparisons considered the total scores of the participants in pre/post and delayed tests. The comparison addressed the following research question:

1. Do different glossing modes (L2 definition alone, L2 definition and audio glossing or L2 definition and video/animation glossing) have any significant impact on L2 learners' vocabulary learning and short and long-term word retention?

This research question includes the three sub-questions of *1 (1) to 1 (3)*, which will be addressed subsequently.

Pre-test in-sum to Immediate post-test in-sum (Between-participant Comparison)

1 (1) *Does glossing have any significant impact on L2 learners' vocabulary learning for short-term word retention?*

To investigate the impact of glossing on L2 learners' vocabulary learning in terms of short-term word retention, the performance of the four groups (control & three experimental groups) was compared from the pre-test in-sum to the immediate post-test in-sum. The purpose of this comparison was similar to PR analysis [43]. Figure 4.3 as well as Table 4.22 show the trend of changes in the vocabulary mean scores of each group across the test sessions from the pre-test in-sum to the delayed post-test in-sum. The means of each group are out of 33 (in-sum).

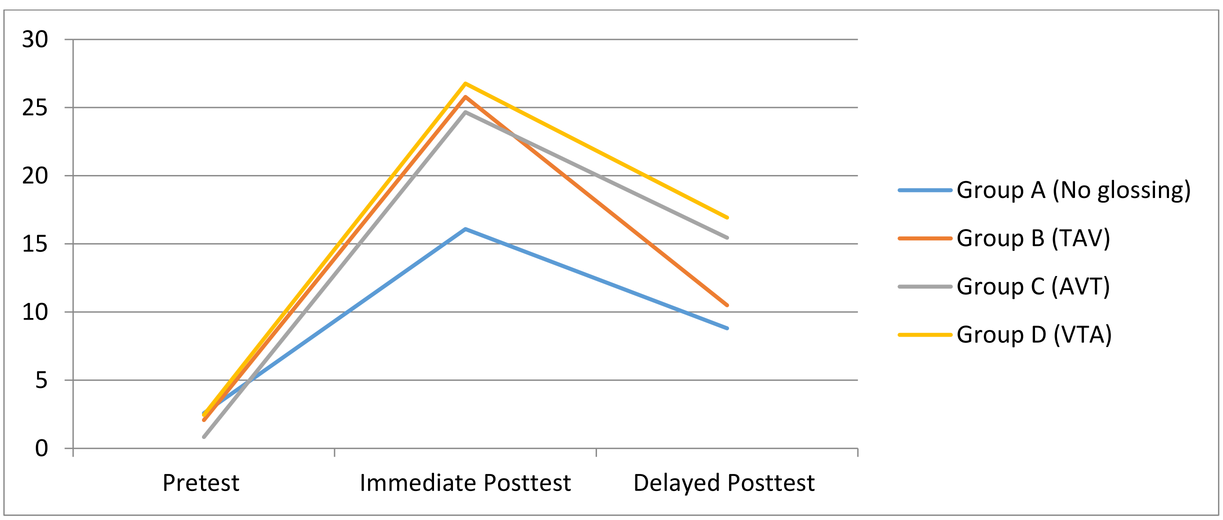


Figure 4. 3 Trend of changes in groups across session from pre-test to delayed post-test in sum (MC test)

[43] Review the discussion of in-sum comparison for this analysis in section one.

Table 4. 22 Descriptive Statistics of 4 Groups from Pre-test to Delayed post-test in-sum (MC* test)

Group		N	Mean	Std. Deviation	Skewness		Kurtosis	
						Std. Error		Std. Error
Gr. A (Control)	Pre-test	24	2.58	3.00	2.608	.472	8.747	.918
	Immediate post-test	22	16.09	4.28	-.955	.491	.576	.953
	Delayed post-test	20	8.80	4.06	.591	.512	.968	.992
Gr. B (TAV)	Pre-test	39	2.07	1.62	.409	.378	-.807	.741
	Immediate post-test	28	25.78	3.70	-.528	.441	-.303	.858
	Delayed post-test	28	10.50	8.32	.278	.383	-.792	.750
Gr. C (AVT)	Pre-test	36	.83	1.27	2.339	.393	6.848	.768
	Immediate. post-test	27	24.66	4.25	-.655	.448	-.053	.872
	Delayed. post-test	27	15.44	6.25	.365	.448	.275	.872
Gr. D (VTA)	Pre-test	33	2.45	2.04	1.429	.409	2.35	.798
	Immediate post-test	30	26.76	3.88	-.492	.427	-.38	.833
	Delayed post-test	28	16.92	6.84	.216	.441	-1.04	.858

*MC stands for multiple-choice productive recognition test.

As the means of the groups (A, B, C, & D) in Table 4.22 show, the performance of the four groups on the pre-test was very low, indicating that the groups had a minimum baseline knowledge regarding the target glossed words before the instruction; yet, with the instruction, the three experimental groups (B, C & D) who received the new words via different glossing modes performed better than group A (control), because they received higher mean scores compared to the control group. Furthermore, group A (control) got the lowest mean score (M= 16.09) in comparison to the three experimental groups. Also, among the three experimental groups, group D (VTA) had the highest mean score (M= 26.76) on the immediate post-tests in-sum; group B (TAV) got the second highest mean score (M= 25.78), and group C (AVT) received the lowest mean

(M= 24.66). Furthermore, the descriptive table and visual graph show that the performance of the four groups (i.e., A, B, C, & D) declined from the immediate post-tests to the delayed post-test according to the groups' vocabulary mean scores in the MC test. Groups D (M=16.92) and A (M=8.80) had the highest and lowest mean scores respectively on the delayed post-tests in-sum. However, the findings in Table 4.22 were only based on the raw scores (i.e., 33 test items), and were descriptively interpreted. In order to statistically measure the groups' changes across the two test sessions (pre/post immediate-test in-sum), and to ascertain that the four groups were significantly different from each other on the immediate post-tests, ANCOVA was utilized; and the scores of the immediate post-tests in-sum were considered as one dependent variable (DV), the four groups formed one independent variable (IV), and the initial pre-test scores acted as the covariate. Before conducting ANCOVA, the assumption of homogeneity of variance was ensured via Levene's test ($p > .05$) (See Appendix R – Table R.7). Table 4.23 depicts the main ANCOVA result for MC vocabulary test in regard to short-term word retention, comparing pre-test scores in-sum to the immediate post-test scores in-sum.

Table 4. 23 Tests of Between-participant Effects: ANCOVA (Pre-test in-sum to Immediate test in-sum) (MC* Test)

Dependent variable: immediate post-test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1721.48[a]	4	430.37	26.69	.000	.51
Intercept	29914.35	1	29914.35	1855.35	.000	.94
Group*pre-test	30.09	3	10.03	.61	.607	.01
Pre-test	19.33	1	19.33	1.19	.276	.01
Group	1626.28	3	542.09	33.62	.000	.49
Error	1644.56	102	16.12			
Total	63899.00	107				
Corrected Total	3366.05	106				

a. R Squared = .511 (Adjusted R Squared = .492)
*Multiple-choice productive recognition test.

The third row in Table 4.23 indicates that the assumption of homogeneity of regression slopes was met ($p > .05$). The fourth row (Pre-test) shows that the 4 groups were not significantly different from each other on the pre-test (i.e., covariate) ($p > .05$). Finally, the fifth row (Group) compares groups A (control), B (TAV), C (AVT), and D (VTA) in terms of their immediate post-tests in-sum. The findings demonstrated that the groups' means were significantly different from each other on the immediate post-tests in-sum ($p < .05$) with a large effect size[44] ($\eta p^2 = .49$), taking into account the covariate.

In order to see where exactly the significant differences lay, post hoc pairwise comparison was conducted among the groups, using Sidak adjustment test for multiple comparisons (See Table 4.24).

Table 4. 24 Pairwise Comparison (Pre-test in-sum to Immediate post-test in-sum) (MC Test)

Dependent variable: immediate post-test. Multiple-choice (MC) productive recognition test

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.^{b}	95% Confidence Interval for Difference^b Lower Bound	Upper Bound
Gr. A (Control)	Gr. B (TAV)	-9.533*	1.153	.000	-12.628	-6.438
	Gr. C (AVT)	-8.200*	1.203	.000	-11.428	-4.972
	Gr. D (VTA)	-10.589*	1.130	.000	-13.621	-7.558
Gr. B (TAV)	Gr. A (Control)	9.533*	1.153	.000	6.438	12.628
	Gr. C (AVT)	1.333	1.101	.789	-1.620	4.286
	Gr. D (VTA)	-1.056	1.057	.901	-3.893	1.781
Gr. C (AVT)	Gr. A (Control)	8.200*	1.203	.000	4.972	11.428
	Gr. B (TAV)	-1.333	1.101	.789	-4.286	1.620
	Gr. D (VTA)	-2.389	1.097	.176	-5.333	.555
Gr. D (VTA)	Gr. A (Control)	10.589*	1.130	.000	7.558	13.621
	Gr. B (TAV)	1.056	1.057	.901	-1.781	3.893
	Gr. C (AVT)	2.389	1.097	.176	-.555	5.333

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

[44] Review section one for the discussion of effect size.

As shown in Table 4.24, groups B (TAV), C (AVT), and D (VTA) were significantly different from group A (control) in their immediate post-test score with larger mean differences ($p < .05$), meaning the performance of the three experimental groups significantly increased on the immediate post-test in comparison to their performance on the pre-test ($p < .05$); and they also performed significantly better than group A (control) on the same test. The finding also signifies that the vocabulary technique of glossing was significantly more effective than the non-glossing strategy for L2 learners' short-term word retention. Additionally, the experimental groups (B, C, & D) were not significantly different from each other on the immediate post-test in-sum ($p > .05$), meaning glossing technique was similarly effective for L2 learners' short-term word learning and recollection. Table 4.25 presents the adjusted means of the four groups in the immediate post-test in-sum after controlling the initial pre-test differences (i.e., covariate).

Table 4. 25 Adjusted Means after Controlling the Covariate (Pre-test in-sum to Immediate test in-sum) (MC* test)

Dependent variable: immediate post-test

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Gr. A (Control)	16.25[a]	.86	14.52	17.97
Gr. B (TAV)	25.78[a]	.75	24.28	27.29
Gr. C (AVT)	24.45[a]	.79	22.87	26.03
Gr. D (VTA)	26.84[a]	.73	25.38	28.30

a. Covariates appearing in the model are evaluated at the following values: Pretest = 2.0374.
*Multiple-choice productive recognition test.

According to Table 4.25, group A (control) has the lowest mean score (M=16.25) in the immediate post-test in comparison to the other three experimental groups; group C (AVT) received the lowest mean (M=24.45), group B (TAV) received the second lowest mean (*M*=25.78), and group D (VTA) achieved the highest mean (*M*=26.84) in the immediate post-test in-sum.

Overall, preliminary checks were conducted to ensure that there was no violation of the assumptions of ANCOVA. After adjusting for the initial pre-test differences, and according to Table 4.23, *and in line with the findings of PR vocabulary test for this comparison*, there was a statistically significant difference among the four groups in the immediate post-test score in-sum, $F3 = 33.62$, $p = .000$, $\eta p^2 = .49$, meaning the vocabulary technique of glossing was significantly more effective than the non-glossing vocabulary strategy for participants' short-term word retention in the experimental groups when compared to the control group. Besides, there was no significant difference among the means of the three experimental groups ($p > .05$), meaning groups B (TAV), C (AVT), and D (VTA) performed similarly on their immediate post-test in-sum.

Immediate post-test in-sum to Delayed post-test in-sum (Within-participant Comparison)

1 (b) *How does glossing affect L2 learners' vocabulary learning in short-term versus long-term?*

The purpose of this comparison was to investigate if the target glossed words were retained from short-term to long-term (i.e., immediate post-test to delayed post-test); and if yes, which gloss order was more effective for this short versus long-term word retention. A within-participant paired samples t-test was conducted, and each group's mean score was compared from the immediate post-test in-sum to the delayed post-test in-sum. As shown earlier in Table 4.22, group A (control) was the only group who had the lowest mean score on the delayed post-test (*M*= 8.80) in comparison to the three experimental groups. Also, the performance of groups B (TAV), C (AVT), and D

(VTA) declined from the immediate post-test to the delayed post-test. The declines were, however, statistically measured for any significant differences. Table 4.26 shows the descriptive statistics of the paired samples (the means of groups are out of 33); and Table 4.27 presents the main findings of the paired samples t-test.

Table 4. 26 Paired Samples Descriptive Statistics (Immediate test in-sum to Delayed test in-sum). Multiple-choice (MC) productive recognition Test

Group			Mean	N	Std. Deviation	Std. Error Mean
Gr. A (Control)	Pair 1	Immediate post-test	15.95	20	4.47	1.00
		Delayed post-test	8.80	20	4.06	.90
Gr. B (TAV)	Pair 1	Immediate post-test	25.78	28	3.70	.70
		Delayed post-test	13.53	28	6.73	1.27
Gr. C (AVT)	Pair 1	Immediate post-test	24.24	25	4.11	.82
		Delayed post-test	15.36	25	6.49	1.29
Gr. D (VTA)	Pair 1	Immediate post-test	26.22	27	3.68	.70
		Delayed post-test	17.00	27	6.96	1.33

According to Table 4.26, the mean scores of groups A (control), B (TAV), C (AVT), and D (VTA) on the delayed post-test in-sum were lower than those of the immediate post-test in-sum. The findings in Table 4.27 also indicate that there was a significant difference between the scores of each group from the immediate post-test in-sum to the delayed post-test in-sum ($p < .05$).

Table 4. 27 Paired Samples t-test (Immediate test in-sum to Delayed test in-sum, mean out of 33). Multiple-choice (MC) productive recognition test

Group			Paired Differences					t	df	Sig. (2-tailed)
			Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
						Lower	Upper			
Gr. A (Control)	Pair 1	Immediate post-test- Delayed post-test	7.15	5.53	1.23	4.55	9.74	5.77	19	.000
Gr. B (TAV)	Pair 1	Immediate post-test- Delayed post-test	12.25	7.50	1.41	9.33	15.16	8.63	27	.000
Gr. C (AVT)	Pair 1	Immediate post-test- Delayed post-test	8.88	8.07	1.61	5.54	12.21	5.49	24	.000
Gr. D (VTA)	Pair 1	Immediate post-test- Delayed post-test	9.22	8.10	1.55	6.01	12.42	5.91	26	.000

A detailed analysis on the paired samples t-test (Table 4.27) reveals that there was a statistically significant decrease in the vocabulary scores of group A (control) from the immediate post-test in-sum (M = 15.95, SD = 4.47) to the delayed post-test in-sum (M=8.80, SD=4.06), t (19) =5.77, p < .000 (two-tailed). The mean decrease in the paired comparison was 7.15 with a 95% confidence interval ranging from 4.55 to 9.74. In group B (TAV), too, a statistically significant decrease was observed in their vocabulary scores from the immediate post-test in-sum (M=25.78, SD=3.70) to the delayed post-test in-sum (M=13.53, SD=6.73), t (27) = 8.63, p < .000 (two-tailed), and the mean decrease was 12.25 with a 95% confidence interval ranging from 9.33 to 15.16. Group C (AVT) showed the same statistically significant decline in their vocabulary scores from the immediate post-test in-sum (M=24.24, SD=4.11) to the delayed post-test in-sum (M =15.36, SD =6.49) to, t (24) = 5.49, p < .000 (two-tailed), and the mean decrease in the paired comparison was 8.88 with a 95% confidence interval ranging from 5.54 to 12.21. Likewise, group D (VTA) revealed a statistically significant drop in the vocabulary scores from the immediate post-test in-sum (*M*=26.22, *SD*=3.68) to the delayed post-test in-sum (*M*=17.00, *SD*=6.96), t (26) = 5.91, p < .000 (two-tailed) with the mean decrease of 9.22 in the paired comparison, and a 95% confidence interval ranging from 6.01 to 12.42.

Therefore, it can be concluded that, *similar to PR vocabulary test results for this comparison*, the participants in three experimental groups (B, C & D) failed to retain a significant number of the target glossed words in long-term as compared to the short-term, and revealed a significant attrition of the learned words in the delayed vocabulary test; meaning the glossing strategy did not significantly contribute to the word retention of the L2 learners in long-term in comparison to their short-term. Yet, unlike the PR test, among the experimental groups, the participants in group B (i.e., TAV) showed more attrition of the learned words in delayed test, t (27) = 8.63, p < .05, d= .7 (large effect size) than the other groups; whereas group C participants (i.e., AVT) revealed less attrition of the learned words in the same test, t (24) = 5.49, p < .05, d = .5 (large effect size) (See Table 4.27).

Pre-test in-sum to Delayed post-test in-sum (Within-participant Comparison)

1 (3) *Does glossing have any significant impact on L2 learners' vocabulary learning for long-term word retention?*

The purpose of this comparison was to examine L2 learners' vocabulary learning and word retention in the long-term. To do so, the pre-test score in-sum was compared to the delayed post-test score in-sum without including the immediate post-test score in-sum. Each group was compared separately from each other (i.e., within-participant comparison) across the two test sessions (pre-and delayed). Table 4.28 presents the descriptive statistics of this comparison, showing the mean score of each group on the delayed post-test in-sum. The means of each group are out of 33.

Table 4. 28 Descriptive Statistics (Comparing Pre-test in-sum to Delayed post-test in-sum) (MC Test)

Dependent variable: delayed post-test. Multiple-choice (MC) productive recognition test

Group	Mean (pre-test)	Mean (delayed)	Std. Deviation	N
Gr. A (Control)	2.58	8.80	4.06	20
Gr. B (TAV)	2.07	10.50	8.32	38
Gr. C (AVT)	.83	15.44	6.25	27
Gr. D (VTA)	2.45	16.92	6.84	28

Group A (control) achieved the lowest mean (M=8.80) in comparison to groups B (TAV), C (AVT), and D (VTA) on the delayed post-test in-sum. Also, among the three experimental groups, group D received the highest mean (M=16.92); and group B received the lowest mean score on the delayed post-test (M=10.50). However, the mean differences in the descriptive table were statistically measured via ANCOVA to ascertain if the four groups were significantly different from one another on the delayed post-test in-sum. In ANCOVA, the effect of the initial pre-test scores in-sum formed one covariate, the delayed post-test scores in-sum were considered as the DV, and the groups acted as one IV. The normality was already checked (See Table 4.26), and the homogeneity of variance was controlled via Levene's test (See Appendix R – Table R.8). The result showed that the assumption was not observed for this analysis ($p < .05$); thus, like similar occasions, the alpha level was set at .025 (Tabachnick & Fidelle, 2013). Table 4.29 demonstrates the main ANCOVA result in regard to long-term word retention.

Table 4. 29 Tests of Between-participant Effects: ANCOVA (Pre-test in-sum to Delayed test in-sum) (MC Test)

Dependent variable: delayed post-test. Multiple-choice (MC) productive recognition test				
Source	df	F	Sig.	Partial eta squared
Corrected model	4	6.19	.000	.18
Intercept	1	183.54	.000	.63
Group*pre-test	3	.07	.976	
Pre-test	1	.00	.963	.000
Group	3	8.20	.000	.18
Error	108			
Total	113			
Corrected total	112			

a. R Squared = .187 (Adjusted R Squared = .157)

Similar to previous ANCOVA tables in this chapter, the third row in Table 4.29 (Group*pre-test) ensured the assumption of homogeneity of regression slopes ($p > .05$); and the fourth row (Pre-test) shows that the four groups were not significantly different from each other on the pre-test (i.e., covariate) ($p > .05$). Finally, the fifth row (Group) compares the control and three experimental groups in terms of their delayed post-tests. The result showed that groups A (control), B (TAV), C (AVT), and D (VTA) were significantly different from each other on the delayed post-test in-sum ($p < .05$) with a large effect size ($\eta p^2 = .18$), taking into account the initial pre-test differences (i.e., covariate). However, to locate where the significant differences lay, a post hoc pairwise comparison was conducted, and the four groups were compared together in terms of their sum of the delayed post-test score (see Table 4.30).

Table 4. 30 Pairwise Comparison (Pre-test in-sum to Delayed post-test in-sum) (MC Test)

Dependent variable: delayed post-test. Multiple-choice (MC) productive recognition test			
(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-1.71	.941
	Gr. C (AVT)	-6.67*	.013
	Gr. D (VTA)	-8.13*	.001
Gr. B (TAV)	Gr. A (Control)	1.71	.941
	Gr. C (AVT)	-4.96*	.036
	Gr. D (VTA)	-6.42*	.002
Gr. C (AVT)	Gr. A (Control)	6.67*	.013
	Gr. B (TAV)	4.96*	.036
	Gr. D (VTA)	-1.46	.972
Gr. D (VTA)	Gr. A (Control)	8.13*	.001
	Gr. B (TAV)	6.42*	.002
	Gr. C (AVT)	1.46	.972

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

The findings indicate that there were significant differences between groups C (AVT) and A (control), C and B (TAV), D (VTA) and A, and D and B ($p < .05$); also, group B (TAV) was not significantly different from group A (control) ($p > .05$). In addition, among the three experimental groups, groups C and D performed significantly better than group B ($p < .05$); meaning the participants in these two groups showed significant long-term word retention than the participants of group B on the delayed vocabulary test. The groups can also be ranked from the lowest performance to the highest performance on the delayed post-test as groups A (control), B (TAV), C (AVT) and D (VTA).

Table 4.31 presents the adjusted means of the four groups in the delayed post-test score in-sum after considering the effect of the covariate (i.e., initial pre-test differences). The finding displays that group A (control) received the lowest mean score (M= 8.78) on the delayed post-test in comparison to the three experimental groups; and the performance of group D (M=16.92) was better than groups B (TAV) and C (AVT).

Table 4. 31 Adjusted Means after Controlling Covariate (Pre-test in-sum to Delayed post-test in-sum) (MC Test)

Dependent variable: delayed post-test. Multiple-choice (MC) productive recognition test		
Group	Mean	Std. error
Gr. A (Control)	8.78[a]	1.56
Gr. B (TAV)	10.50[a]	1.12
Gr. C (AVT)	15.46[a]	1.37
Gr. D (VTA)	16.92[a]	1.31

a. Covariates appearing in the model are evaluated at the following values: Pretest = 2.0708.

Altogether, after preliminary checks on the assumptions of ANCOVA, and adjusting the means for initial pre-test differences, according to Table 4.29, *and in line with PR vocabulary test results*, the results showed that there was a statistically significant difference among the four groups in the delayed post-test scores in-sum, $F3$ = 8.20, p = .000, ηp^2= .18 (large effect size). However, it can be concluded, with caution, that that the glossing technique of vocabulary instruction was partially effective for L2 learners' long-term word retention; because groups C and D only showed significant differences on their delayed vocabulary test in comparison to groups A and B ($p < .05$).

It should be noted that the above-mentioned comparisons considered only the sum of the pre/immediate and delayed post-test scores. Also, the four groups were compared separately from one another; and the sub pre/immediate and delayed post-tests were not included. In other words, the comparisons did not specify which glossing mode(s) assisted L2 learners to learn the target glossed words significantly better than another in terms of both short and long-term word retention. To resolve this issue, detailed between-participant comparisons, including sub-test comparisons, was carried out. The next section presents the findings.

Sub-test Comparisons

Sub-test comparisons considered participants' sub-test scores on each test. As stated before, in order to analyze the sub-test comparisons, the sub/immediate and delayed post-tests were created. The sub-test comparisons addressed the following research question:

2. *Do different glossing modes (L2 definition alone, L2 definition and audio glossing or L2 definition and video/animation glossing) have any significant impact on L2 learners' vocabulary learning and short and long-term word retention?*

This question includes two sub research questions (1.1 & 1.3), which are analyzed and interpreted in the following sections. Also, another analysis was done to compare participants' scores from sub immediate to sub delayed post-tests for short versus long-term effect (i.e., 1.2). Table 4.32 presents the descriptive statistics of the four groups (the control & three experimental groups), showing their performance from the sub pre-test to the sub delayed post-test across test sessions. The means represent each group's performance on the sub-tests.

Table 4. 32 Descriptive Statistics of Groups from Sub pre-tests to Sub delayed post-tests (MC* test)

Group		N	Mean	Std. deviation	Skewness		Kurtosis	
						Std. error		Std. error
Gr. A (Control)	Pre.Sub1	24	1.50	2.91	3.74	.47	15.70	.918
	Pre.Sub2	24	.79	.88	.85	.47	-.04	.918
	Pre.Sub3	24	.2917	.62	2.06	.47	3.23	.918
	Im.Post.Sub1	24	4.70	2.27	.03	.47	-.57	.918
	Im.Post.Sub2	24	4.91	1.79	.38	.47	-.66	.918
	Im.post.Sub3	22	6.50	2.57	-1.16	.49	.86	.953
	Del.Post.Sub1	20	3.45	2.23	1.56	.51	2.69	.992
	Del.Post.Sub2	20	2.47	1.63	-.25	.50	-1.45	.972

	Del.Post.Sub3	20	2.75	1.55	.65	.51	.34	.992
Gr. B (TAV)	Pre.Sub1	39	.76	1.11	1.94	.37	4.56	.741
	Pre.Sub2	39	.82	.75	.31	.37	-1.15	.741
	Pre.Sub3	39	.48	.68	1.09	.37	.00	.741
	Im.Post.Sub*1	30	9.30	2.56	-1.01	.42	.98	.833
	Im.Post.Sub2	30	7.83	2.50	-.75	.42	-.38	.833
	Im.Post.Sub3	31	8.54	1.67	-.82	.42	-.83	.821
	Del.Post.Sub*1	28	3.53	3.08	.22	.37	-1.28	.741
	Del.Post.Sub2	28	3.28	2.83	.51	.38	-.55	.750
	Del.Post.Sub3	28	3.48	2.90	.24	.37	-1.11	.741
Gr. C (AVT)	Pre.Sub1	36	.138	.42	3.27	.39	10.99	.768
	Pre.Sub2	36	.583	.76	1.29	.39	1.44	.768
	Pre.Sub3	36	.11	.39	3.87	.39	15.55	.768
	Im.Post.Sub1	33	8.81	2.42	-1.31	.40	2.78	.798
	Im.Post.Sub2	30	7.56	2.17	-.68	.42	1.36	.833
	Im.Post.Sub3	30	8.20	2.26	-1.02	.42	-.26	.833
	Del.Post.Sub1	27	4.92	2.60	-.05	.44	-.71	.872
	Del.Post.Sub2	27	5.48	2.53	.69	.44	-.28	.872
	Del.Post.Sub3	27	5.03	2.44	.33	.44	-.64	.872
Gr. D (VTA)	Pre.Sub1	33	1.00	1.06	1.00	.40	.62	.798
	Pre.Sub2	33	1.03	1.13	.90	.40	.00	.798
	Pre.Sub3	33	.42	.86	2.69	.40	8.51	.798
	Im.Post.Sub1	32	9.68	2.38	-.78	.41	-.42	.809
	Im.Post.Sub2	31	8.48	1.76	-.41	.42	-.23	.821
	Im.Post.Sub3	31	8.77	1.45	-.97	.42	-.10	.821
	Del.Post.Sub1	28	6.42	3.27	.02	.44	-.95	.858
	Del.Post.Sub2	28	5.07	2.38	-.26	.44	-.89	.858
	Del.Post.Sub3	28	5.27	2.34	-.28	.43	-1.09	.845

* MC stands for multiple-choice productive recognition test.
*Note: Im. Post. Sub stands for sub immediate post-test and Del. Post. Sub stands for sub delayed post-test.
*Number of test items: sub-pre/immediate & post-test 1 = 12 items; sub-pre/immediate & delayed 2 = 11 items; &
sub-pre/immediate & delayed 3 = 10 items.

According to Table 4.32, group A (control) achieved the lowest mean scores on both sub immediate and delayed post-tests in comparison to groups B (TAV), C (AVT), and D (VTA). The three experimental groups (B, C & D) also performed better than group A in all the 3 sub immediate post-tests; however, their performance declined on the corresponding sub delayed post-tests. The next section presents the inferential findings regarding the four groups' changes across sessions (i.e., between-participant comparisons); and examines if the mean differences were statistically significant among the four groups from sub pre-tests to sub delayed post-tests. The between-participant comparisons also investigate which glossing mode(s) was significantly effective for L2 learners' short and long-term word learning and retention. All the relevant assumptions for sub-test analyses were also checked and ensured.

Sub pre-test to Sub immediate post-test (Between-participant Comparison)

2 (1) *Which glossing mode (s) contributes significantly to L2 learners' short-term word learning and retention?*

To do the analysis, all four groups (control & three experimental groups) were compared together from the sub pre-tests to the corresponding sub immediate post-tests, using ANCOVA with the effect of the sub pre-test scores as the covariate; sub immediate post-test scores as one DV, and groups as one IV. The assumption of homogeneity of variance was ensured via Levene's Test ($p > .05$) (See Appendix R – Table R.9). Table 4.33 presents the main findings of the three ANCOVAs, comparing each sub pre-test to the relevant sub immediate post-test across the four groups (A, B, C & D).

Table 4. 33 Tests of Between-participant Effects: ANCOVA (Sub pre-test to Sub immediate post-test) (MC Test)

Dependent variable: immediate post-test. Multiple-choice (MC) productive recognition test

	Comparing pre1-im1				Comparing pre2-im2				Comparing pre3-im3			
Source	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared
Corrected Model	4	17.40	.000	.37	4	11.77	.000	.30	4	5.04	.001	.15
Intercept	1	1018.82	.000	.89	1	823.46	.000	.88	1	1475.04	.000	.93
Group* Pre.sub1	3	.05	.984		3	.58	.626		3	3.18	.027	
Pre.sub1	1	.25	.615	.00	1	2.9	.088	.02	1	.77	.380	.00
Group	3	21.90	.000	.36	3	15.0	.000	.29	3	6.59	.000	.15
Error	114				110				109			
Total	119				115				114			
Corrected Total	118				114				113			

The assumption of homogeneity of regression slopes in the third row was met for sub comparisons 1 and 2 ($p > .05$), but violated for sub-comparison 3 ($p < .05$). However, this violation could not cause any problem in the subsequent interpretation of the results pertinent to this analysis, because, according to Hamilton (1977), ANCOVA is robust against this violation when sample sizes do not vary a lot. The fourth row in Table 4.33 (Pre. sub 1) shows that the four groups were not significantly different from each other in the pre-test (i.e., covariate) ($p > .05$). Finally, the fifth row (Group) compares the groups (control & three experimental groups) in terms of their sub immediate post-tests. The findings show that groups A (control), B (TAV), C (AVT), and D (VTA) were significantly different from each other on the sub immediate post-tests ($p < .05$) with large effect sizes (ranging from .15 to .36). However, to examine where exactly the significant differences lay, post hoc pairwise comparison was conducted, and the four groups were compared together in terms of their sub immediate post-tests (See Table 4.34).

Table 4. 34 Pairwise Comparison (Sub pre-test to Sub immediate post-test) (MC Test)

Dependent variable: Immediate post-tests 1, 2, & 3. Multiple-choice (MC) productive recognition test

		Comparing pre1-im1		Comparing pre2-im2		Comparing pre3-im3	
(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-4.538*	.000	-2.932*	.000	-2.089*	.002
	Gr. C (AVT)	-4.011*	.000	-2.578*	.000	-1.655*	.023
	Gr. D (VTA)	-4.943*	.000	-3.633*	.000	-2.299*	.000
Gr. B (TAV)	Gr. A (Control)	4.538*	.000	2.932*	.000	2.089*	.002
	Gr. C (AVT)	.527	.952	.354	.986	.434	.956
	Gr. D (VTA)	-.405	.987	-.701	.719	-.210	.999
Gr. C (AVT)	Gr. A (Control)	4.011*	.000	2.578*	.000	1.655*	.023
	Gr. B (TAV)	-.527	.952	-.354	.986	-.434	.956
	Gr. D (VTA)	-.931	.575	-1.056	.275	-.644	.766
Gr. D (VTA)	Gr. A (Control)	4.943*	.000	3.633*	.000	2.299*	.000
	Gr. B (TAV)	.405	.987	.701	.719	.210	.999
	Gr. C (AVT)	.931	.575	1.056	.275	.644	.766

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

The pairwise comparison revealed that group A (control) had significantly lower mean differences in all the sub immediate post-tests ($p > .05$), and the three experimental groups achieved significantly higher means than the control group ($p < .05$); meaning groups B, C, and D outperformed group A in their performance on all the immediate post-tests. Also, there was no significant difference among the performance of the three experimental groups (B, C, & D) in sub immediate post-tests ($p > .05$), meaning groups B (TAV), C (AVT), and D (VTA) performed similarly on sub immediate post-tests. Thus, *unlike PR results*, the findings of Table 4.34 show that different glossing modes were equally effective for L2 learners' short-term word retention.

Table 4.35 presents the adjusted means of the 4 groups (control & three experimental groups) on sub immediate post-tests while controlling the covariate. As shown, group A (control) had the lowest mean score in all the sub immediate post-tests in comparison to groups B (TAV), C (AVT), and D (VTA). Besides, the means of groups B, C and D were almost similar on the three sub immediate post-tests.

Table 4. 35 Adjusted Means after Controlling the Covariate Effect (Sub pre-test to Sub immediate test) (MC Test)

Dependent variables: immediate post-tests 1,2, & 3. Multiple-choice (MC) productive recognition test

	Comparing pre1-im1		Comparing pre2-im2		Comparing pre3-im3	
Group	Mean	Std. error	Mean	Std. error	Mean	Std. error
Gr. A (Control)	4.759[a]	.506	4.914	.424	6.494	.425
Gr. B (TAV)	9.297[a]	.444	7.846	.379	8.583	.360
Group C (AVT)	8.770[a]	.434	7.491	.382	8.149	.368
Group D (VTA)	9.702[a]	.431	8.547	.375	8.793	.358

In sum, according to between-participant analysis (Table 4.33), the findings show that there were statistically significant differences from sub pre-test 1 to sub immediate post-test 1, $F3$ = 21.90, p = .000, ηp^2= .36 (large effect size); from sub pre-test 2 to sub immediate post-test 2, $F3$ = 15.0, p = .000, ηp^2= .29 (large effect size); and finally, from sub pre-test 3 to sub immediate post-test 3, $F3$ = 6.59, p = .000, ηp^2= .15 (large effect size); yet, since the three experimental groups were not significantly different from one another, it can be concluded that different glossing modes were equally effective for L2 learners' word learning and retention in short-term.

Sub immediate post-test to Sub delayed post-tests (Between-participant Comparison)

2 (2) *Which glossing mode (s) affect L2 learners' vocabulary learning in short-term versus long-term?*

In order to investigate which glossing mode(s) led to less attrition of the target glossed words in comparison to the short-term word retention, the four groups were compared from sub immediate post-tests to the corresponding sub delayed post-tests.

Figure 4.4 displays the visual line graph of the four groups in each sub-test from sub immediate to sub delayed post-test.

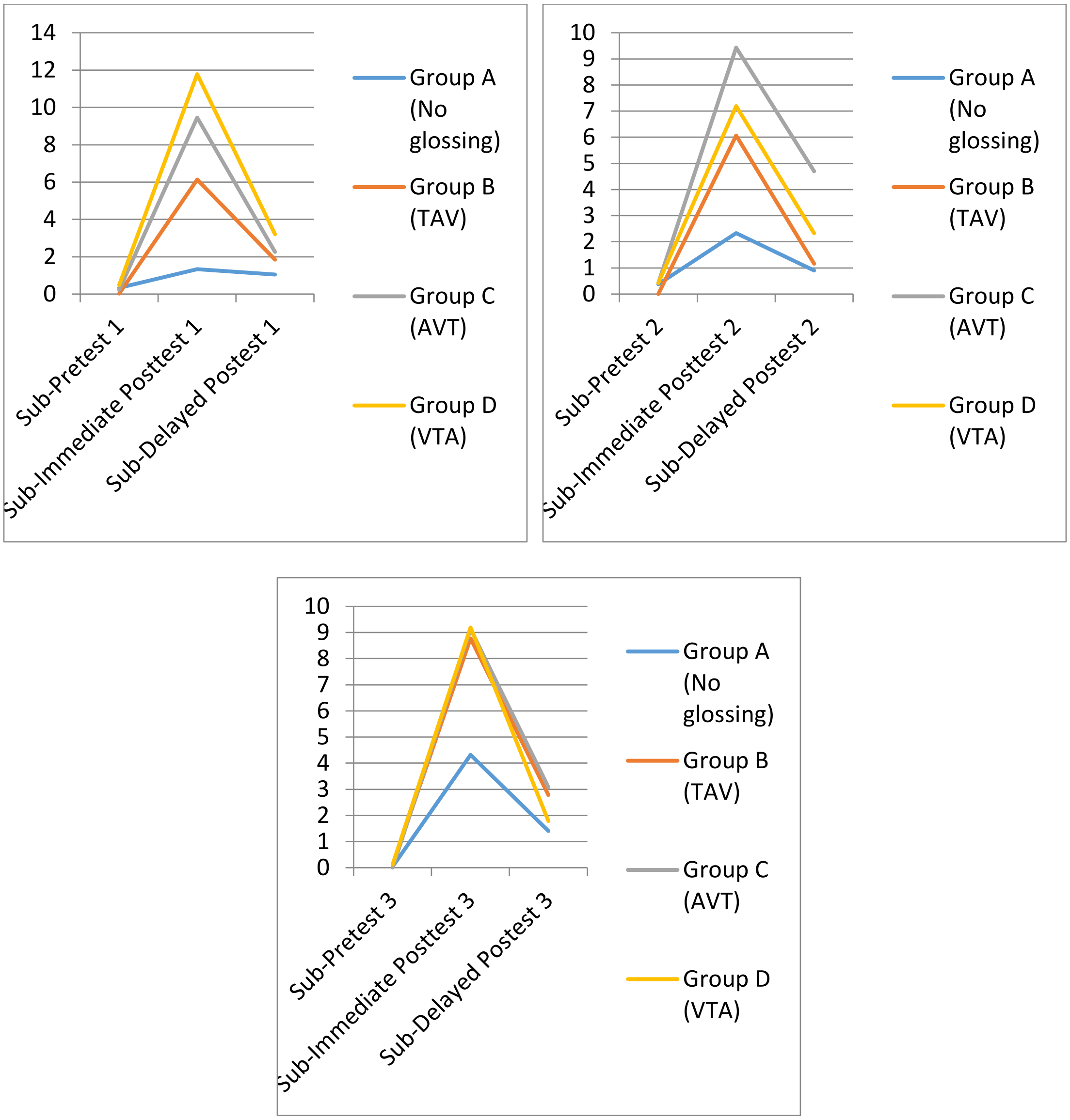


Figure 4. 4 Trend of group changes across test sessions (MC vocabulary test)

The graph shows that groups declined from sub immediate to sub delayed post-tests. Group A (control) performed the lowest in comparison to groups B (TAV), C (AVT), and D (VTA) on the delayed vocabulary tests in comparison to their immediate post-tests. The inferential test of ANCOVA was conducted to compare the scores from sub immediate to the corresponding sub delayed post-tests across the groups while taking into account the effect of the covariate (i.e., sub immediate post-test differences). The assumptions of ANCOVA were checked and observed (See Appendix R – Table R.10). Table 4.36 presents the main results of three ANCOVAs for this comparison.

Table 4. 36 Tests of Between-participant Effects: ANCOVA (Sub immediate to Sub delayed tests across Groups)

	Comparing Im1-del1				Comparing Im2-del2				Comparing Im3-del3			
Source	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared
Corrected Model	4	3.869	.006	.134	4	5.701	.000	.189	4	4.114	.004	.139
Intercept	1	35.575	.000	.262	1	12.622	.001	.114	1	20.082	.000	.164
Group* Im.post1	3	1.148	.334		3	1.383	.253		3	2.345	.078	
Im.post1	1	1.130	.290	.011	1	1.529	.219	.015	1	.018	.893	.000
Group	3	4.887	.003	.128	3	4.361	.006	.118	3	4.860	.003	.125
Error	100				98				102			
Total	105				103				107			
Corrected Total	104				102				106			

The third row confirms that the assumption of the homogeneity of regression slopes was assumed ($p > .05$). The fourth row (Im. post 1) indicates that the four groups (control & three experimental groups) were not significantly different from each other on the sub immediate post-tests (i.e., covariate) ($p > .05$); and the fifth row (Group) compared groups A (control), B (TAV), C (AVT), and D (VTA) in terms of their sub delayed vocabulary tests. As shown, groups A, B, C, and D were significantly different from each other in their performance on all sub delayed post-tests ($p < .05$) with medium effect sizes (ranging from .11 to .12). However, to locate the significant differences among the groups, post hoc pairwise comparison was conducted (See Table 4.37). This comparison also identifies which glossing mode(s) resulted in less attrition of the learned glossed words from short to long-term.

Table 4. 37 Pairwise Comparison (Sub immediate post-tests to Sub delayed post-tests across Groups) (MC Test)

Dependent variable: immediate post-tests 1, 2, & 3. Multiple-choice (MC) productive recognition test

		Comparing Im1-del1		Comparing Im2-del2		Comparing Im3-del3	
(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-1.73	.381	-1.06	.678	-1.34	.316
	Gr. C (AVT)	-1.99	.210	-2.55*	.009	-2.26*	.012
	Gr. D (VTA)	-3.57*	.003	-2.12	.062	-2.49*	.005
Gr. B (TAV)	Gr. A (Control)	1.73	.381	1.06	.678	1.34	.316
	Gr. C (AVT)	-.26	1.000	-1.48	.139	-.91	.600
	Gr. D (VTA)	-1.84	.075	-1.05	.478	-1.14	.315
Gr. C (AVT)	Gr. A (Control)	1.99	.210	2.55*	.009	2.26*	.012
	Gr. B (TAV)	.26	1.000	1.48	.139	.91	.600
	Gr. D (VTA)	-1.57	.206	.43	.988	-.22	.999
Gr. D (VTA)	Gr. A (Control)	3.57*	.003	2.12	.062	2.49*	.005
	Gr. B (TAV)	1.84	.075	1.05	.478	1.14	.315
	Gr. C (AVT)	1.57	.206	-.43	.988	.22	.999

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

In pair comparison between sub immediate and delayed post-tests 1, the performance of the participants in group A (control) was not significantly different from those in groups B (TAV) and C (AVT) ($p > .05$); but significantly different from group D (VTA) ($p < .05$). In other words, L2 learners in group D performed significantly better than the learners in group A on sub delayed post-test 1, and retained the learned words from short-term (sub immediate post-test 1) to long-term (sub delayed post-test 1) (i.e., less attrition of the words in long-term as compared to the short-term), $F3 = 4.88$, $p = .003$, $\eta p^2 = .12$ (medium effect size). The participants in group D received the target glossed words via the dual mode of L2 definition and video/animation glossing on day 1. Hence, this glossing mode assisted them to retain the words better from short-term to long-term than the glossing modes of L2 definition and audio glossing or L2 definition alone. Also, the three experimental groups were not significantly different from each other ($p > .05$). Therefore, the findings should be interpreted with caution for day 1.

In the second comparison, the participants in group C (AVT) performed significantly different from those in group A (control) ($p < .05$), but not from groups B (TAV) and D (VTA) ($p > .05$), $F3 = 4.36$, $p = .006$, $\eta p^2 = .11$ (medium effect size). The participants in group C were instructed through the dual mode of L2 definition and video/animation glossing on day 2; thus, this dual glossing mode supported the retention of the learned words better from short-term to long-term than the glossing modes of L2 definition and audio glossing or L2 definition alone. Also, the three experimental groups were not significantly different from each other ($p > .05$). Thus, the finding here should again be concluded with caution.

Finally, the third sub-test comparison showed that the participants in groups C (AVT) and D (VTA) were the only ones who showed a significant difference on sub delayed vocabulary test 3 in comparison to groups A (control) and B (TAV) ($p < .05$), $F3 = 4.86$, $p = .003$, $\eta p^2 = .12$ (medium effect size); however, the three experimental groups were not significantly different from each other ($p > .05$); hence the results should be interpreted with caution. L2 learners in groups C and D showed less attrition of the learned words from short-term to the long-term in comparison to groups A and B only. The participants in group C and D received the new words via the two glossing modes of L2 definition alone and L2 definition and audio glossing respectively on day 3.

Table 4.38 presents the adjusted means of the four groups on sub delayed post-tests after considering the effect of the covariate. As shown, group A (control) received the lowest mean on all sub delayed vocabulary tests in comparison to groups B (TAV), C

(AVT), and D (VTA); and the mean scores of groups B, C, and D were also low compared to their immediate post-test test scores.

Table 4. 38 Adjusted Means after Controlling Covariate (Sub immediate test to sub delayed test) (MC Test)

Dependent variable: immediate post-tests 1, 2 & 3. Multiple-choice (MC) productive recognition test						
	Comparing Im1-del1		Comparing Im2-del2		Comparing Im3-del3	
Group	Mean	Std. error	Mean	Std. error	Mean	Std. error
Gr. A (Control)	2.99[a]	.753	2.84[a]	.602	2.77[a]	.554
Gr. B (TAV)	4.72[a]	.519	3.91[a]	.444	4.12[a]	.424
Gr. C (AVT)	4.98[a]	.536	5.39[a]	.482	5.03[a]	.450
Gr. D (VTA)	6.56[a]	.538	4.96[a]	.477	5.26[a]	.440

As a result, the between-participant comparisons (Table 4.36) show that the four groups declined significantly from sub immediate to sub delayed post-tests with medium effect size ($p < .05$). Yet, in order to examine which glossing mode (s) helped L2 learners to experience less attrition from short to long-term (sub immediate to sub delayed post-tests), Table 4.37 reveals that on days 1 and 2 of the instruction, the dual mode of L2 definition and video/animation glossing contributed significantly to L2 learners' less attrition of the target glossed words from short-term to long-term; and the two glossing modes of L2 definition alone, and L2 definition and video/animation glossing were effective for the participants' less word attrition from short to long-term on day 3. The next section analyzes the findings in regard to long-term word learning and retention across the four groups.

Sub pre-test to Sub delayed post-test (Between-group Comparison)

2 (3) *Which glossing mode (s) contributes significantly to L2 learners' long-term word learning and retention?*

The purpose of this comparison was to investigate which glossing mode (s) was significantly more effective for the long-term word learning and retention of L2 learners across groups (control & three experimental groups). Table 4.39 presents the descriptive statistics of this comparison, including the means of the four groups on sub delayed vocabulary tests.

Table 4. 39 Descriptive Statistics (Sub pre-test to Sub delayed post-test without Sub-immediate) (MC Test)

Dependent variable: sub delayed post-test. Multiple-choice (MC) productive recognition test

	Sub-test 1			Sub-test 2			Sub-test 3		
Group	Mean	Std. deviation	N	Mean	Std. deviation	N	Mean	Std. deviation	N
Gr. A (Control)	3.45	2.23	20	2.47	1.63	21	2.75	1.55	20
Gr. B (TAV)	3.53	3.08	39	3.28	2.83	38	3.48	2.90	39
Gr. C (AVT)	4.92	2.60	27	5.48	2.53	27	5.03	2.44	27
Gr. D (VTA)	6.42	3.27	28	5.07	2.38	28	5.27	2.34	29

According to Table 4.39, group A (control) received the lowest mean scores on all sub delayed post-tests, as compared to groups B (TAV), C (AVT), and D (VTA). Besides, group D on sub delayed post-tests 1 and 3 (M= 6.42; M= 5.27 respectively), and group C on sub delayed post-test 2 (M= 5.48) received the highest mean scores in comparison to the other groups (A & B). It is interesting to note that, according to Table 4.32 above, participants' mean scores in all four groups (A, B, C, & D) increased from sub pre-tests to sub delayed post-tests. However, an inferential test of ANCOVA was conducted to examine if the mean differences among the groups were statistically significant; it compared the participants' scores on each sub pre-test to the corresponding sub delayed post-test without including the sub immediate post-tests across groups. In this analysis, groups formed one IV, the sub delayed test scores formed one DV, and the initial pre-test scores acted as one covariate. All ANCOVA assumptions were checked; however, the result of Levene's test showed the violation of this assumption for sub delayed post-test 3 (See Appendix R – Table R.11); thus, the alpha level was set at .025

(Tabachnick & Fidelle, 2013). Table 4.40 demonstrates the findings of the three ANCOVAs for this comparison.

Table 4. 40 Tests of Between-participant Effects: ANCOVA (Sub pre-tests to Sub delayed post-tests) (MC Test)

Dependent variable: sub delayed post-test. Multiple-choice (MC) productive recognition test

	Sub-test 1				Sub-test 2				Sub-test 3			
Source	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared	df	F	Sig.	Partial eta squared
Corrected Model	4	4.934	.001	.153	4	6.669	.000		4	4.671	.002	
Intercept	1	199.298	.000	.646	1	172.517	.000		1	245.526	.000	
Group * Pre.sub1	3	.090	.966		3	.820	.485		3	.753	.523	
Pre.sub1	1	.011	.916	.000	1	.800	.373		1	.081	.776	
Group	3	6.578	.000	.153	3	8.635	.000	.19	3	6.123	.001	.14
Error	109				109				110			
Total	114				114				115			
Corrected Total	113				113				114			

The homogeneity of regression slopes was also assured (p > .05) (third row in Table 4.40). The fourth row (Pre.sub1) shows that the 4 groups (control & three experimental groups) were not significantly different from each other on all sub delayed post-tests (p > .05). Finally, the fifth row (Group) compared the 4 groups in terms of their sub delayed post-tests. As shown, groups A (control), B (TAV), C (AVT), and D (VTA) were significantly different from each other on sub delayed post-tests (p < .05) with large effect size (ranging from .14 to .19), considering the covariate. Post hoc pairwise comparison was conducted to examine where the significant differences existed among the four groups (See Table 4.41).

Table 4. 41 Pairwise Comparison (Sub pre-tests to Sub delayed post-tests) (MC Test)

Dependent variable: sub delayed post-test. Multiple-choice (MC) productive recognition test

		Sub-test 1		Sub-test 2		Sub-test 3	
(I) Group	(J) Group	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]	Mean difference (I-J)	Sig.[b]
Gr. A (Control)	Gr. B (TAV)	-.10	1.000	-.8	.783	-.75	.854
	Gr. C (AVT)	-1.50	.454	-2.98*	.000	-2.26*	.015
	Gr. D (VTA)	-2.98*	.004	-2.641*	.002	-2.53*	.004
Gr. B (TAV)	Gr. A (Control)	.10	1.000	.82	.783	.75	.854
	Gr. C (AVT)	-1.39	.309	-2.15*	.005	-1.51	.102
	Gr. D (VTA)	-2.88*	.001	-1.82*	.023	-1.78*	.024
Gr. C (AVT)	Gr. A (Control)	1.50	.454	2.98*	.000	2.26*	.015
	Gr. B (TAV)	1.39	.309	2.15*	.005	1.51	.102
	Gr. D (VTA)	-1.48	.337	.33	.997	-.26	.999
Gr. D (VTA)	Gr. A (Control)	2.98*	.004	2.64*	.002	2.53*	.004
	Gr. B (TAV)	2.88*	.001	1.82*	.023	1.78*	.024
	Gr. C (AVT)	1.48	.337	-.33	.997	.26	.999

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Sidak.

Table 4.41 shows that, on the first sub delayed post-test, the performance of groups A (control), B (TAV) and C (AVT) were not significantly different from each other ($p > .05$). Also, group D (VTA) performed significantly better than groups A (control) and B (TAV) ($p < .05$), but not significantly better than group C (AVT) ($p > .05$), meaning the participants in group D could retain the words in long-term better in comparison to those in groups A and B ($F3 = 6.57$, $p = .000$, $\eta p^2 = .15$). However, group D participants did not differ from group C participants in the same test. The participants in group D received the target glossed words via the dual mode of L2 definition and video/animation glossing on day 1; thus, this mode helped the learners to retrieve the words in long-term better than the two glossing modes of L2 definition and audio glossing, and L2 definition alone, or even no glossing mode.

On the second sub delayed post-test, the performance of the participants in groups A (control) and B (TAV); and C (AVT) and D (VTA) were not significantly different from each other ($p > .05$), but the two groups of C and D were significantly different from groups B and A ($p < .05$). In other words, L2 learners in groups C and D retained the learned words in long-term significantly better than the other participants in group A and B ($F3 = 8.63$, $p = .000$, $\eta p^2 = .19$). The two glossing modes of L2 definition and video/animation, and L2 definition alone were used to instruct the target words on day 2.

Finally, on sub delayed test 3, there was no significant difference between groups A (control) and B (TAV) ($p > .05$); B and C (AVT); and C and D (VTA) ($p > .05$). However, the participants in groups D and C performed significantly better than those in group A (control) on the sub delayed post-test 3 ($p < .05$); and only group D participants (with L2 definition and audio glossing mode) performed significantly better than groups A (with no glossing mode) and B (with L2 definition and video/animation glossing mode) ($p < .05$), and retrieved the learned words in long-term ($F3 = 6.12$, $p = .001$, $\eta p^2 = .14$). In other words, the two modes of L2 definition and audio glossing as well as L2 definition alone had a significantly positive effect on L2 learners' long-term word learning and retention on day 3.

Table 4.42 displays the adjusted means of the four groups (control & three experimental groups) on sub delayed tests after considering the covariate effect. As shown, group A (control) achieved the lowest mean on all sub delayed post-tests; and group D (VTA) on sub delayed tests 1 ($M = 6.42$) and 3 ($M = 5.28$), and group C (AVT) on sub delayed test 2 ($M = 5.01$) received the highest mean scores.

Table 4. 42 Adjusted Means after Controlling Covariate (Sub pre-tests to Sub immediate post-tests) (MC Test)

Dependent variable: sub delayed post-test. Multiple-choice (MC) productive recognition test

	Sub-test 1		Sub-test 2		Sub-test 3	
Group	Mean	Std. error	Mean	Std. error	Mean	Std. error
Gr. A (Control)	3.43[a]	.66	2.46[a]	.54	2.74[a]	.55
Gr. B (TAV)	3.54[a]	.46	3.28[a]	.40	3.49[a]	.39
Gr. C (AVT)	4.93[a]	.57	5.44[a]	.47	5.01[a]	.48
Gr. D (VTA)	6.42[a]	.55	5.10[a]	.46	5.28[a]	.46

Overall, the between-participant comparison (Table 4.40) shows that the control and three experimental groups were significantly different from each other on all sub delayed post-tests ($p < .05$). However, to examine which glossing mode (s) caused a significantly positive impact on L2 learners' long-term word learning and retention, the pairwise comparison (Table 4.41) reveals that the dual mode of L2 definition and video/animation glossing was effective for the participants' vocabulary retention in long-term on day 1; the two glossing modes of L2 definition and video/animation as well as L2 definition alone were effective for L2 learners' long-term retention on day 2; and the two modes of L2 definition and audio glossing as well as L2 definition alone positively influenced L2 learners' word retention in long-term on day 3.

Summary of Section Two

In section two, the findings of multiple-choice productive recognition vocabulary test were presented and analyzed. Two types of between and within-participant comparisons were carried out via ANCOVA and paired samples t-test respectively. The data were once analyzed with in-sum comparisons, and another time with sub-test comparisons.

The results of both between and within in-sum comparisons showed that: **(a)** the vocabulary technique of glossing was significantly more effective than the non-glossing strategy for participants' short-term word learning and retention; **(b)** the glossing technique did not contribute significantly to L2 learners' vocabulary retention in long-term when compared to the short-term. However, among the experimental groups, the participants in group B (i.e., TAV) showed more attrition of the learned words in delayed test; whereas group C participants (i.e., AVT) revealed less attrition of the learned words in the same test; and finally, **(c)** the vocabulary strategy of glossing was partially effective for L2 learners' long-term word learning and retention.

The findings of between sub-test comparisons revealed that: **(a)** the three glossing modes were equally effective for L2 learners' short-term word learning and retention; **(b)** glossing modes were differently effective for L2 learners' attrition of the target words from short-term to long-term word retention; on days 1 and 2 of the instruction, the dual mode of L2 definition and video/animation glossing contributed to L2 learners' less attrition of the target glossed words from short to long-term; and the two glossing modes of L2 definition alone, and L2 definition and video/animation glossing were effective for participants' attrition of the learned words from short-term to long-term on day 3; and finally, **(c)** different glossing modes contributed significantly to participants' word learning and retention in long-term. The dual mode of L2 definition and video/animation glossing was effective for L2 learners' long-term word learning on day 1; the two glossing modes of L2 definition and video/animation as well as L2 definition alone were effective for learners' word retention in long-term on day 2; and the two modes of L2 definition and audio glossing as well as L2 definition alone positively influence participants' word retention in long-term on day 3.

SECTION THREE

(Questionnaire & Semi-structured Interview)

1. *What are L2 learners' attitudes and perceptions towards simultaneous multimedia glossing? and which glossing mode (s) do they prefer, and why?*

The next section presents and interprets the findings of the questionnaire, followed by the semi-structured interviews.

Questionnaire

Eighty-three participants from the three experimental groups (i.e., B, C, & D) filled out the questionnaire. They were asked to respond to a set of 15 close and open-ended questions. The first 12 close-ended questions were scored on a 5-point Likert-type scale, with 1 being strongly agree and 5 being strongly disagree for questions 1-9; and 1 being extremely helpful and 5 being unhelpful for questions 10-12[45]. The close-ended questions asked participants' opinions about the type of glossing modes in this study. Questions 1, 4, 7, and 10 asked if participants found the glossing mode of *L2 definition alone* an easy way to learn and remember new words, and whether or not they would use it for their future vocabulary learning. Following the same themes, questions 2, 5, 8, and 11 reflected on learners' perception about *L2 definition and audio glossing mode*; and questions 3, 6, 9, and 12 sought learners' viewpoints regarding *L2 definition and*

[45] See the discussion on the questionnaire for 5-point Likert scale in Methodology chapter.

video/animation glossing. The participants were also required to rank their choice of glossing modes from 1 to 3, with 1 being their first option, and 3 being the last choice. Finally, the two open-ended questions (14 & 15) required the participants to state their rationale for their choice of glossing mode(s), and add their comments regarding the use of other vocabulary techniques for word learning and retention. Tables 4.43-4.46 present the frequency[46] of responses for questions 1-12, along with the mean of each question in the questionnaire. The relevant analyses (i.e., questions 1-3, 4-6, 7-9, and 10-12) are shown in the following tables respectively, and the report is presented after each table.

Questions 1-3 of the questionnaire asked the participants to state if it was easy for them to learn the new words with text-definition alone, text-definition and audio pronunciation, or text-definition and video/animation glossing respectively (See Table 4.43).

Table 4. 43 Frequency of Responses to Questions 1-3

	Frequency (Q.1: Text- definition alone)	Frequency (Q.2: Text-definition & audio)	Frequency (Q.3: Text-definition & video/animation)
Strongly agree (1)	7 (8.43%)	18(21.68%)	59 (71%)
Agree (2)	27 (32.53%)	45(54.21%)	19 (23%)*
Neither/nor (3)	29 (34.93%)	13(15.66%)	4 (4.81%)*
Disagree (4)	14 (16.86%)	6 (7.22%)	0 (0%)
Strongly disagree (5)	6 (7.22%)	1 (1.2%)	1 (1.2%)
Mean	2.81	2.12	1.37

Note: * percentages are rounded for the ease of reporting.

According to Table 4.43, findings show that: **(a)** 94% of the participants (71% + 23% which constitute 78 out of 83 participants) strongly agree or agree that it was easy for them to learn the target glossed words via text/L2 definition and video/animation

[46] To calculate the frequency of response in percentage, the given number for each question is divided by the total number of the participants, multiplied by 100.

glossing (question 3, *M*= 1.37). The mean shows that most of the responses on the Likert scale lean towards 1 (strongly agree) or 2 (agree). Only 1.2 % of the participants strongly disagree to the effectiveness of video/animation glossing mode for an easier vocabulary learning trial; **(b)** approximately 76 % of the participants (63 participants) strongly agree or agree that it was easy for them to learn the new words via L2 definition and audio glossing mode (question 2, *M* = 2.12) (i.e., 54.21% agree, and 21.68 % strongly agree). The mean score of question 2 confirms that the learners tended to choose 1 (strongly agree) or 2 (agree) on the Likert scale; and finally, **(c)** approximately 35% of the participants (29 out of 83) had indifferent views about whether or not the single glossing mode was an easier way for them to learn new words in comparison to the dual glossing modes. Thus, looking at the overall responses, it could be implied that the mean of responses to this question was close to number 3 in the Likert scale (*neither agree nor disagree*), meaning the participants gave neutral reply to the question. Additionally, only less than 10 % of the participants (7 out of 83) *strongly agree* and almost 33 % (27 out of 83) *agree* that learning new words would be facilitated if accompanied by L2 definition alone (question 1, *M* = 2.81); and nearly 24% of the participants (20 out of 83) *strongly disagree* and *disagree* together that it was not easy for them to acquire the target glossed words via L2 definition alone. In sum, when comparing the three glossing modes in terms of how easy it was for the participants to learn the new words, the dual glossing mode of L2 definition and video/animation comes first, the bimodal glossing of L2 definition and audio comes second, and the single glossing mode of L2 definition comes last.

Questions 4-6 of the questionnaire asked the participants if it was easier for them to remember the target words in the final test when receiving the instruction via text-definition alone, text-definition and audio pronunciation, or text-definition and video/animation glossing (See Table 4.44).

Table 4. 44 Frequency of Responses to Questions 4-6

	Frequency (Q.4: Text-definition alone)	Frequency (Q.5: Text-definition & audio)	Frequency (Q.6: Text-definition & video/animation)
Strongly agree	7 (8.43%)	12(14.45%)	42(50.60%)
Agree	19 (22.89%)	41(49.39%)	29(34.93%)
Neither/nor	25 (30.12%)	20(24.09%)	7(8.43%)
Disagree	24 (28.91%)	8(9.63%)	3 (3.61%)
Strongly disagree	8 (9.63%)	2 (2.4%)	1 (1.2%)
Mean	3.08	2.36	1.66

With regards to Table 4.44, it was revealed that: **(a)** more than half of the participants[47] (42% or 50.60%) *strongly agree* or *agree* (29% or 34.93%) that the incorporation of video/animation glossing with L2 definition assisted them to remember the new words better in the final vocabulary tests (question 6). The mean (M = 1.66) also confirms that the participants selected 1 (strongly agree) for this question; **(b)** nearly 64% of the participants (53 out of 83) *strongly agree* or *agree* that the glossing mode of text-definition and audio recording helped them to remember the words easier than other glossing modes in the final vocabulary tests (question 5, M = 2.38). Only 24% of the learners (20 out of 83) had an indifferent opinion (i.e., *neither agree* nor *disagree*) regarding this glossing mode, and around 11 % of the students (10 people) *disagree* or *strongly disagree* about the efficacy of audio glossing presentation for their long-term

[47] The percentage of those participants who strongly agree (i.e., 50% or 42 out of 83) and those who agree (i.e., 35% or 29 out of 83) are reported together.

word retention; **(c)** in addition, 26 participants (about 31% or 7 +19=26 learners) *strongly agree* or *agree* that target glossed words could be remembered easier when instructed via L2 definition alone, 25 people (around 30%) *neither agree nor disagree* on this choice, and only 32 participants (38.55% or 24 + 8 learners) *strongly disagree* or *disagree* about the easiness of remembering new words through L2 definition alone glossing (question 4, M = 3.08). The average of the responses to this question confirmed the frequency of responses regarding the Likert scale 3 (*neither agree nor disagree)*.

Questions 7-9 asked participants about their willingness to use the glossing modes of L2 definition alone, L2 definition and audio glossing, and L2 definition and video/animation glossing as their three possible vocabulary learning strategies/choices in future while learning a new word. Table 4.45 presents the findings.

Table 4. 45 Frequency of Response to Question 7-9

	Frequency (Q.7: Text-definition alone)	Frequency (Q. 8: Text-definition & audio)	Frequency (Q. 9: Text-definition & video/animation)
Strongly agree	11 (13.25%)	14 (16.86%)	54 (65%)
Agree	19 (22.89%)	39 (46.98%)	22 (26.5%)
Neither/nor	26 (31.32%)	23 (27.71%)	5 (6%)
Disagree	17 (20.48)	4 (4.81%)	1 (1.2%)
Strongly disagree	10 (12%)	3 (3.61%)	1 (1.2%)
Mean	2.95	2.31	1.46

As the findings show, it was revealed that: **(a)** 65% of the participants (54 out of 83 *strongly agree*) as well as 26.5% (22 learners *agree*) believed that they were willing to use the glossing technique of text-definition and video/animation in future for their vocabulary learning (question 9, M = 1.46); however, very few participants either *disagree* (1.2%) or *strongly disagree* (1.2%) to employ this vocabulary strategy in future, and only 5 participants (6%) neither *agree* nor *disagree* on the use of video/animation mode for their future word learning; **(b)** about 64% of the participants (39 + 14 = 53 out of 83) *strongly agree* or *agree* that they will be using the vocabulary technique of L2 definition and audio glossing for their future word learning (question 8, M = 2.31), meaning the participants' average was 2 (agree) from the Likert scale. However, 27.71% of the participants (23 out of 83) neither *agree* nor *disagree* to employ this mode in future. It is also interesting to note that in total, nearly 8.5% of the participants (7 out of 83) either *strongly disagree* or *disagree* to practice word learning via L2 definition and audio glossing mode; and finally, **(c)** given the choice, among 30 participants who responded to *strongly agree* (1) or *agree* (2) option in the scale, 11 participants (13.25%) *strongly agree* to select L2 definition alone for their future vocabulary learning, and the other 19 participants (22.89%) just *agree* to use this strategy for later use. Also, around 31% had a neutral idea regarding the use of this glossing strategy for later use, and 27 participants (32.53%) *strongly disagree* or *disagree* with the choice of learning new words through L2 definition mode in future (question 7, M = 2.95). The mean of responses to this question was close to number 3 in the Likert scale (*neither agree nor disagree*), meaning the majority of the participants gave a neutral reply to this question.

Table 4.46 shows the results of questions 10-12 which asked the participants if they found the three glossing modes (i.e., L2 definition alone, L2 definition & audio glossing, and L2 definition & video/animation glossing) extremely helpful or helpful, somewhat helpful, neither helpful nor unhelpful, or unhelpful.

Table 4. 46 Frequency of Responses to Questions 10-12

	Frequency (Q.10: Text-definition alone)	Frequency (Q.11: Text-definition & audio)	Frequency (Q.12: Text-definition & video/animation)
Extremely helpful	8 (9.63%)	11 (13.25%)	65 (78.31%)
Helpful	25 (30.12%)	47 (56.62%)	13 (15.66%)
Somewhat helpful	31 (37.34%)	20 (24.09%)	4 (4.81%)
Neither/nor	16 (19.27%)	3 (3.61%)	1 (1.2%)
Unhelpful	3 (3.61%)	2 (2.4%)	0 (0%)
Mean	2.77	2.25	1.28

The findings in Table 4.46 revealed that: **(a)** altogether, nearly 94% of the learners (78 out of 83: 65 +13 participants) found the vocabulary strategy of L2 definition and video/animation glossing *extremely helpful* or *helpful* for their word learning, meaning the participants had a positive attitude towards using this multimedia-based vocabulary strategy (question 12, M = 1.28). In other words, the questions that asked L2 learners' opinions about the incorporation of L2 definition and video/animation glossing for word learning and long-term retention had a mean score of 1.28, which leans towards the rating scale of 1 (*extremely helpful*) to 2 (*helpful*) on the Likert scale. It is worth mentioning that no one referred to this display mode as *unhelpful* (0%); **(b)** more than half of the learners (i.e., nearly 70% or 47 +11 participants) regarded the display mode of L2 definition and audio glossing *extremely helpful* or *helpful* for word learning, and 20 participants (24%) viewed it as *somewhat helpful* (question 11, M = 2.25). The mean score of this question leans between 2 and 3, meaning participants had generally a positive perception towards the dual mode of L2 definition and audio glossing. About 4% of the participants had neutral views (i.e., *neither helpful nor unhelpful*) about this display condition glossing format, and only 2% of the learners (2 out of 83) found the combination of L2 definition and audio glossing *unhelpful*; and finally, **(c)** in terms of the glossing mode of L2 definition alone, the data indicated that less than 10% of the participants (8 people) found single glossing mode *extremely helpful*, 30% (25 people) viewed it as *helpful*, and around 37% of the learners (31 out of 83) rated it as *somewhat helpful*. Also, less than 5% of the learners found this glossing mode *unhelpful* (question 10, M = 2.77). As implied, the mean of this response is close to 3, which means that L2 learners tended to select number 3 (*somewhat helpful*) more in the Likert scale. Hence, the findings suggest that the glossing mode of L2 definition alone was not completely a preferred mode of vocabulary instruction for L2 learners' word learning and long-term retention in the present study. Also, this glossing mode seemed to be their least favorite choice for future vocabulary learning endeavours compared to L2 definition and audio glossing, and/or L2 definition and video/animation glossing.

Table 4.47 presents the analysis of question 13 of the questionnaire. The participants were asked to rank the three glossing modes (i.e., L2 definition alone; L2 definition & audio glossing; and L2 definition & video/animation glossing) in the order of their preferences, with 1 as their first priority and 3 as their last favorite choice.

Table 4. 47 Frequency of Response to Question 13

Codes*	Frequency	Valid Percent
1-2-3 (TAV)	3	3.61
1-3-2 (TVA)	5	6.02
2-1-3 (ATV)	2	2.40
2-3-1 (AVT)	11	13.25
3-1-2 (VTA)	6	7.22
3-2-1 (VAT)	56	67.46*

* Code meanings: 1 → Participants' preference/priority to L2 definition alone glossing (*T*)
2 → Participants' preference/priority to L2 definition and audio glossing (*A*)
3 → Participants' preference/priority to L2 definition and video/animation glossing (*V*)
* Percentages are rounded for the ease of reporting.

As the findings show, the highest frequency belonged to 3-2-1 (i.e., VTA), meaning nearly 68% of participants (56 out of 83 participants) preferred L2 definition and video/animation glossing first, L2 definition and audio glossing second, and L2 definition alone last; 11 participants (13.25%) selected L2 definition and audio glossing first, L2 definition and video/animation glossing second, and L2 definition alone last (i.e., AVT); and very few participants (3 out of 83 or 3.61%) chose the L2 definition alone first, L2 definition and audio second, and L2 definition and video/animation last (i.e., TAV). Another interpretation of Table 4.47 could be that the presentation of target glossed words via L2 definition and video/animation glossing mode was the most preferred format of vocabulary learning for only approximately 75% of the participants (67.46% + 7.22%), meaning the rest of the 25% of the learners either chose L2 definition along with audio glossing mode as their first choice (around 16% of the learners:13.25% + 2.40%) or selected L2 definition alone as their most favorite means of word acquisition (9% of the participants: 3.61% + 6.02%). Thus, it can be implied that even though learning new words through the simultaneous presentation of L2 definition and video/animation glossing has been the most favorable vocabulary learning technique for the majority of L2 learners in the present study, only a small number of the participants liked to listen to the voice of a native-speaker, pronouncing the new words and its definitions as their first option; and very few learners perceived L2 definition alone as the most effective strategy for their vocabulary learning and retention.

In Question 14 of the questionnaire, the participants were asked about their rationale for the choice of glossing strategies that they preferred the most in question 13. Some participants justified their reasons based on the fact that learning words via watching a relevant video/animation clip helped them to remember the words for later use; others commented that using this mode could assist them to learn new words sooner, and they could guess them in context better; few participants mentioned the practicality of using video/animation mode for the correct word pronunciation. Furthermore, those who chose the dual mode of L2 definition and audio glossing for word learning believed that the strategy was effective for pronunciation improvement. Finally, among the participants who adhered to the glossing mode of L2 definition alone, some stated that this mode helped them to learn new words easily, and improved their vocabulary reservoir due to expanding word knowledge by learning more synonyms.

Question 15 asked L2 learners' viewpoints about any other vocabulary learning strategies they used before for word learning. Based on participants' mostly cited responses, some vocabulary learning strategies such as providing examples for new words, learning them in groups and categories, acquiring words by watching video/animation clips or movies, utilizing new words in real-life and practical/actual situations, and listening to the pronunciations of the new words were considered as the useful vocabulary learning techniques that almost all the participants practiced. As for participants' other choices, the comments that entailed *using games, vocabulary activities/drills/flashcards/pictures/symbols/taking-note techniques and vocabulary quizzes* were the most frequently stated themes. It is also interesting to note that learners' *preference to use dictionary* and *first language (L1) equivalents* for word learning, as well as *word repetition, overlearning technique,* and *jotting the words down* were among other frequently-cited comments/statements.

In sum, the findings of the questionnaire showed that the participants perceived learning new words via the dual glossing mode of L2 definition and video/animation as their most favorable vocabulary learning technique. L2 definition and audio glossing was their second mode of preference; whereas, the single glossing mode of L2 definition alone was their last selection. As for L2 learners' rationale and motive to prefer the simultaneous presentation of L2 definition and video/animation glossing over the two other modes, they unanimously agreed that watching videos and animation clips assisted them to learn the words better, and remembered them longer for later use.

Face-to-face Semi-Structured Interview Guide

In order to understand participants' attitudes and preferences towards simultaneous multimedia glosses (i.e., L2 definition alone, L2 definition & audio glossing, and L2 definition & video/animation glossing), the participants were also invited to a 30-minute interview[48]. The interviews were audio-recorded and transcribed, and the common themes were identified[49]. The following section reports the findings.

The participants agreed that the simultaneous presentation of L2 definition and video/animation glossing was the most effective vocabulary technique that assisted them to acquire the words faster and retain them longer. As an example, one interviewee, *Mehran*[50], stated that watching the video/animation clips of the new words along with their simultaneous L2 definitions could help him to function better in the delayed post-

[48] Review methodology chapter for the discussion on participant recruitment.
[49] Review methodology chapter for the discussion on theme extraction.
[50] All the interviewees' names are pseudonyms to protect their privacy.

test, as he could associate the word meaning to the relevant images/animation clips: *"We keep the images of the words in the mind when we learn them by text and video/animation. This strategy helped me a lot to remember the words after two weeks."* To *Sepehr*, another interviewee, vocabulary learning via video/animation mode was an enjoyable learning experience, which motivated him to learn the target words effectively, and save them in his memory for long-term use: *"I enjoyed this learning experience as I was strongly motivated to learn the words, repeat them several times, and use them later."* However, another participant, *Maryam*, favored the vocabulary mode of L2 definition and audio glossing as one effective word learning technique. This aural mode enabled her to learn the correct pronunciation of the new words, and gave her a chance to repeat the new words several times for a better long-term word retention afterwards: *"When I know how a word is pronounced, I can learn it better. The audios of the words helped me to remember the definitions fast."*

In a follow-up question, I asked about the participants' opinion about the form of presenting target glossed words in this study (i.e., simultaneous mode or single mode). The interviewees preferred the concurrent presentation of the glossed words rather than the single mode as it helped them to learn the words in greater depth.

The participants were also asked about the common strategies and techniques they used to learn and remember the new vocabulary. Of the 9 interviewed participants (three from each experimental group), 6 people cited that they tried to learn words in categories while using other techniques such as synonyms and antonyms, word repetitions, watching videos, using games, vocabulary activities, drills, flashcards, taking-note techniques and vocabulary quizzes. Three participants mentioned that they tried to use the new words in different sentences and applied them in communicative contexts such as verbal/oral communication.

Overall, the findings of the interviews indicated that L2 learners perceived the dual mode of L2 definition and video/animation glossing as the most effective vocabulary learning technique for word learning and retention. In addition, the dual mode of L2 definition and audio glossing was not as effective as the video/animation mode, but was more effective than L2 definition alone mode; and the single glossing mode was not effectively helpful for L2 learners' word learning experience and retention.

Summary of Section Three

The results of the questionnaire as well as the semi-structured interview showed that L2 learners preferred the glossing mode of L2 definition and video/animation as their most favorite vocabulary learning technique. Thus, it can be concluded that the dual mode of L2 definition and video/animation glossing helped the participants to learn the target glossed words easier, and retain them better in long-term. Also, this glossing technique was one of the participants' first options when it came to the choice of vocabulary instruction/learning techniques for future use.

Also, the two other modes (i.e., L2 definition & audio glossing, and L2 definition alone) received the second and third most favorable word-learning strategies respectively. From the findings, it can also be implied that the vocabulary technique of L2 definition and audio glossing assisted L2 learners to learn and remember the target glossed words better and easier, and L2 learners found this glossing mode helpful for their future word learning.

The next chapter discusses the findings of the three sections regarding the current literature and two theoretical frameworks of the study.

CHAPTER FIVE: DISCUSSION

Introduction

This study aimed at investigating the effectiveness of simultaneous multimedia glossing on L2 learners' vocabulary learning in terms of both short and long-term word retention, and it drew upon Mayer's (2014, 2007) cognitive theory of multimedia learning and Paivio's (1986) dual-coding theory.

There were four groups (one control & three experimental groups) in this study; the participants of the experimental groups received three reading passages with 33 hyperlinked glossed words in three different glossing modes (i.e., L2 definition alone, L2 definition & audio glossing, and L2 definition & video/animation glossing) during the instructional days, while the control group received no glossing instruction. L2 Learners' understanding of the general vocabulary knowledge as well as their knowledge of the target words were measured via a Vocabulary Levels Test and two vocabulary pre-tests of productive recall (PR) and multiple-choice (MC) productive recognition tests. The two immediate and delayed vocabulary tests were also used to gauge learners' word acquisition and retention in short (immediately after the instruction) and long-term (two weeks later) retention. Data were analyzed using ANCOVA and paired samples t-test with both in-sum and sub-test comparisons for the impact of glossing and different glossing modes respectively[51]. Additionally, participants' perceptions and attitudes

[51] In-sum comparisons addressed research question 1 and its sub-questions; and sub-test comparisons addressed research question 2 and its sub-questions.

towards the three glossing modes of vocabulary instruction were assessed through a 5-point Likert-scale questionnaire and face-to-face semi-structured interviews.

The purpose of this chapter is to discuss the findings of the research in light of the current literature and the two theoretical frameworks of multimedia learning. This chapter is divided into three sections. In the first section, the results regarding the impact of glossing in short and long-term retention (i.e., in-sum test comparisons) are overviewed with relevant supporting and contradictory arguments/discussion. This section addresses research question 1 and its sub-questions (1.1 & 1.2). In the second section, the impact of glossing mode(s) in terms of short and long-term retention (i.e., sub-test comparisons) is discussed under two separate parts: (a) relating the findings to the available literature, and (b) relating the findings to the theoretical frameworks. This section addresses research question 2 and its sub-questions (2.1 & 2.2). In the third section, learners' attitudes and perceptions towards different glossing modes are interpreted. This section also addresses the findings related to research question 3.

Brief Overview of Findings and Discussion

Section One: In-sum Test Comparisons

Impact of Glossing on short-term word Retention: PR & MC Tests

The results of the in-sum test comparisons (i.e., impact of glossing) unanimously showed that the vocabulary technique of glossing was significantly more effective than the non-glossing vocabulary strategy for participants' short-term word retention in both the PR and MC productive recognition vocabulary tests. The results addressed research question 1 and sub-question 1.1. The findings are consistent with a growing body of relevant studies suggesting that glossing provides learners with extra information required for text comprehension, and is thus one practical and prominent technique to enhance vocabulary acquisition and retention (AbuSeileek, 2011, 2008; Akbulut, 2007; Al-Seghayer, 2001; Bowles, 2004; Cheng & Good, 2009; Chen & Yen, 2013; Chun & Plass, 1996a, 1996b; Gettys et al., 2001; Ghahari & Heidarolad, 2015; Hong, 2010; Hulstijn, 1992; Hulstijn et al., 1996; Jacobs et al., 1994; Kim & Kim, 2012; Khezrlou & Ellis, 2017; Knight, 1994; Ko, 2012, 2005; Lyman-Hager & Davis, 1996; Nation, 2001; Paribakht & Wesche, 1996; Rassaei, 2017; Rott, 2005; Rott & Williams, 2003; Salem, 2006; Varol & Erçetin, 2016; Watanabe, 1997; Yoshii, 2014, 2006). The results also support explicit/intentional instruction of vocabulary learning (Sadeghi et al., 2016; Chen & Yen, 2013).

As for the positive effectiveness of glossing, literature has shown that glosses direct learners' attention to new words, and encourages them to process the meanings of unfamiliar words (Yoshii, 2014) through creating consciousness-raising and input enhancement (Ghahari & Heidarolad, 2015; Nagata, 1999). The results of the present study showed that glossing contributed to increased vocabulary learning; and the reason might be due to the fact that when learners encountered unfamiliar glossed words in a reading text, their attention was raised through different annotation types (i.e., glossing conditions), resulting in a better processing of lexical items in mind. Thus, the lexical processing might have increased participants' awareness to the new words and encouraged word learning (Rott, 2007; Nation, 2001).

Another factor that contributes to the effectiveness of glossing and increased word learning is that it can accelerate word recognition and assist learners to allocate part of their working memory[52] capacity for word processing and text comprehension (Varol & Erçetin, 2016; Babaie & Razmjoo, 2015). In addition, the instruction of new words via the glossing technique can provide learners with several exposures to unfamiliar words, and hence increase their word retention (Al-Seghayer, 2001; Chun & Plass, 1996b; Hong, 2010; Ko, 2005; Kost et al., 1999; Lomicka, 1998; Nation, 2001, 2013; Plass & Jones, 2005; Schmitt, 2008; Yoshii, 2006). Stewart and Cross (1993) report that through glossing, there is a likelihood that learners use their previous word knowledge to infer the meaning of the unfamiliar words from the text. Since guessing from context is a challenging task, and might lead to wrong inferences (Nation, 2001), glossing can help to prevent erroneous guessing (Hulstijn, 1992). AbuSeileek (2011) stated that by using glosses, the learners are not left alone to "infer or deduce" the meaning of new words, "a process which may lead to wrong inference, especially for readers with limited language proficiency" (p. 1287). Thus, glossing in this study, aimed to provide the participants with exposure to target words through different gloss modalities so that they can make correct guesses while engaging with the text and increase their word retention.

Studies have shown that the use of glosses can save learners' time while reading a text (Nation, 2001) and caters for different learning preferences (Jacobs et al., 1994). Fostering learner autonomy (Nation, 2002; Rott, Williams, & Cameron, 2002; Stewart & Cross, 1991), and making learners less dependent (Gardener, 2011) are also other advantages of glossing that encourage learners to be responsible for the comprehension of the text through their involvement with the text and the target glossed words. Since class

[52] "Working memory explains how readers process new information to place it in long-term memory (….) and how they search and retrieve the information when they need to remember it" (Varol & Erçetin, 2016, p. 760).

timing is mostly limited to instruct all target words, using glossing in the present research could help the participants to learn how to utilize this practical vocabulary strategy for their future word learning.

In line with the benefits of the glossing technique for word acquisition and reading comprehension, Erçetin (2003) and Sakar and Erçetin (2005) found that glossing was motivational, because the learners had an easy access to the glossed words in different modalities which made “reading authentic texts more manageable and enjoyable” (Chen, 2016, p. 413). In other words, glossing substitutes the dictionary (Yanguas, 2009) and provides greater use of authentic and complicated texts (Nation, 2013; Jacobs et al., 1994). In a similar vein, the participants of the present study accessed the meanings of low frequency target words through three different gloss modalities, which assisted them to comprehend the passages better and retrieve the words longer. In sum, the findings confirmed that glossing is a practical vocabulary strategy to enhance word learning and retention in short-term.

Impact of Glossing on Long-term Word Retention: PR & MC Tests

The results of in-sum test comparisons (i.e., impact of glossing) showed that the glossing technique was partially effective for L2 learners’ long-term word retention in both PR and MC vocabulary tests (from pre- to delayed post-tests). Thus, the findings, which address research question 1 and sub-question 1.2, should be considered with caution. In other words, according to the results, the participants had some gains from the pre-tests to delayed post-tests as their scores had increased, which showed that glossing was effective for long-term word retention. However, this word gain was partial because as far as PR test was concerned, only groups B and C participants performed better than groups A and D in the delayed post-tests; and learners in experimental groups of C and D showed significant differences on their delayed vocabulary test in comparison to groups A and B in MC productive recognition tests.

However, this word gain could be strongly consolidated if L2 learners had enough exposure, input, and focus after the explicit/intentional vocabulary instruction. The exposure could aid them to establish form-meaning connections between the target glossed words and their definitions (Schmitt, 2008). According to Ellis (2006) and Yongqi Gu (2003), a single exposure is quite unlikely to lead to word knowledge gain. Also, Rott and Williams (2003) stated that "sizable word gain may require eight to twelve exposures" (p. 48) to an unknown word; yet, there are studies which found that *two* to *four* encounters are sufficient for a significant vocabulary gain (Rott, 1999; Hulstijn et al., 1996); and four-time (F4) exposure for a long-term word retention (Rott, 2007). As an example, Rott (1999) found that two exposures were adequate to result the small but measurable knowledge of the words, but six encounters led to a "significantly larger gains in both productive and receptive knowledge (p. 187). Comparing the learning of words from L2-L1 word pairs, Webb (2007a) found that "a single context had little effect on gaining word knowledge" (p. 75); nevertheless, to develop full knowledge of a word, more than ten repetitions may be required (Webb, 2007b). This finding is in line with Nation's (2001) well-established saying that a lexical item needs to be met many times in order to be learned. Waring and Takaki (2003) also considered 20 encounters to be necessary for learning form-meaning connection.

To support the prominent role of exposure and context, Schmitt (2008) states that:

> Words will have to be met in many different contexts in order to develop mastery of the different word knowledge types, and this entails a long-term recursive approach to vocabulary learning. (Schmitt, 2008, p. 335)

Schmitt (2010) further explains that engagement and amount of time a learner spent on a vocabulary task can result in a successful word learning and recall. However, regarding the vocabulary task, as Laufer and Rozovski-Roitblat (2015) clarify, it is the type of the task which determines how many encounters or exposures a learner needs to make for successful word learning; "the more demanding the type of knowledge is; the more exposures are needed in the same task type" (p. 708). Thus, including both an explicit and intentional word instruction component, as well as "a component which maximizes repeated exposures to lexical items" (Schmitt, 2008, p. 334), such as incidental instruction, in a vocabulary learning program can help increase learners' word knowledge and long-term recollection. Furthermore, as empirical studies show, only few L2 words can be retained from exposures to reading passages if no subsequent vocabulary practices and drills are followed (Laufer, 2006). Therefore, any vocabulary learning approach can involve implementing several post-reading tasks and activities that focus on target words, leading to better vocabulary retention in long-term (Schmitt, 2008) and promote learners' involvements both cognitively and motivationally (Yoshii, 2014).

Keeping all above in mind, the rationale for why participants of this study revealed partial improvement or increase from pre-to delayed post-tests might be attributed to the small number of exposure or encounters to the target words and a lack of adequate reinforcement and vocabulary practices as well as post-reading tasks after the intervention. Using consolidating techniques such as word-focused activities shortly after the instruction and exposing learners to extensive readings (Paribakht & Wesche, 1996) could help learners to remember the words better in the long-term when required.

Also, some post-reading activities such as reading journals, stories, and book reports (Webb & Chang, 2015) can help enhance long-term word development of the learners.

Additionally, the findings of the present study can be further justified relying on some studies reporting the insignificant impact of glossing strategy on learners' long-term word retention (Jacobs et al., 1994); or the ineffective role of glossing for vocabulary acquisition and reading comprehension (Jung, 2016; Bell & Le Blanc, 2000). Jacobs et al. (1994) note that although glossing was positively more effective than non-glossing for learners' vocabulary acquisition shortly after the instruction, the impact did not last in long-term due to the lack of adequate exposure and reinforcement during the four-week period before the vocabulary re-test. The insignificant effect of glossing for long-term retention was also confirmed by Varol and Erçetin (2016), Watanabe (1992), and Black, Wright, Black and Norman (1992). The use of glosses also affected learners' reading comprehension and word acquisition negatively in other studies (Sakar & Erçetin, 2005; Ariew & Erçetin, 2004).

In sum, glossing is one practical technique that focuses "explicit attention on lexical items during exposure" (Schmitt, 2008, p. 352). It consolidates the form-meaning association, which plays a pivotal role in word retrieval (Rott & William, 2003). Research shows that glossing can act as a vocabulary enhancement tool to help learners infer the meaning of some words that "may not be easy to be inferred or guessed correctly from any of the contexts in which the words appear" (Wang, 2016, p. 300). Additionally, glossing facilitates vocabulary acquisition, develops ease of text comprehension, and encourages further L2 readings. With glossed conditions, learners are able to do more lexical processing, leading them to higher word retention (Arpaci, 2016). Schmitt (2008) indicates that any language learning activities that involve words can improve learner participation and engagement in word use, thus leading to word learning and durable retention of vocabulary. Finally, in line with Yoshi's (2014) findings, learners of this study were able to partially sustain the word gains from the pre- to delayed post-tests, but needed to reinforce the words after being learnt (Jacobs et al., 1994; Yee, 2010; Yoshii, 2014) during the two-week gap, which signifies "the necessity for repetition and exposure to the input" (Arpaci, 2016, p. 25). In other words, factors such as using the target words by the participants after reading (Webb & Chang, 2015), or in speech and writing (Newton, 2013), contribute positively for durable long-term word gains.

Section Two: Sub-test Comparisons

Impact of Glossing Modes on Short-term Word Retention: PR & MC Tests

The findings of the sub-test comparisons (i.e., impact of different glossing modes) showed that, as far as PR vocabulary test was concerned, the dual glossing modes (i.e., L2 definition & audio glossing, and L2 definition & video/animation glossing) were effective for *two* out of *three* test sessions[53] in short-term; and all three glossing modes were equally effective for *one* out of *three* test sessions. Also, the glossing modes were similarly effective for participants' short-term word learning and retention in MC productive recognition tests (See Table 5.1). The findings which address research question 2 and sub-question 2.1 for both tests are consistent with the studies that found

[53] Test sessions refer to the three immediate post-tests after the instructions.

the superiority of dual glossing modes over single glossing condition for word learning and retention in short-term (Al-Seghayer, 2001; Chun & Plass, 1996; Yanguas, 2009).

Table 5. 1 Comparing Glossing Modes in Short-term for PR & MC

PR Test			**MC Test**		
Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
Dual glossing modes (video & audio)	Dual glossing mode (video)	All glossing modes equally effective	All glossing modes equally effective		

Impact of Glossing Modes (Long-term Word Retention): PR & MC Tests

The findings of the sub-test comparisons (i.e., impact of different glossing modes) revealed that the dual glossing modes were effective in *two* out of *three* test sessions[54], and the single glossing mode was only influential in *one* out of *three* test sessions regarding L2 learners' long-term vocabulary retention in the PR test. In contrast, the dual glossing modes were effective for all *three* test sessions of the MC productive recognition test in terms of learners' word retention in the long-term with single glossing mode being positively influential for *two* out of *three* test sessions (See Table 5.2). The findings address research question 2 and sub-question 2.2.

Table 5. 2 Comparing Glossing Modes in Long-term for PR & MC Tests

PR Test			MC Test		
Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
Dual glossing mode (video)	Dual glossing mode (video)	Single glossing mode	Dual glossing mode (video)	Dual (video) & Single glossing modes	Dual (audio) & Single glossing modes

[54] Test sessions refer to the three delayed post-tests two weeks after the instruction.

Putting together the findings of both short and long-term word retention in terms of the effectiveness of glossing modes on learners' vocabulary acquisition, it can be tentatively and cautiously discussed that combining definitions of target words with associated verbal (L2 definition & audio) and visual (L2 definition & video/animation) representations is more effective than providing the textual definition of the words alone in assisting vocabulary learning (Al-Seghayer, 2016, 2000; Chun & Plass, 1996; Lin & Tseng, 2012; Sadeghi et al., 2016; Salem, 2006; Yeh & Wang, 2003; Yoshii & Flaitza, 2002). The facilitative impacts of dual glossing modes were observed in both the PR and MC productive recognition tests where participants performed significantly better, not only on the immediate post-tests, but also partially on the delayed post-tests, meaning the two glossing modes had a positive effect on learners' short and long-term word retention. Having said this, the single glossing mode was also shown to be effective for *1* out of *6* test sessions in PR; and *2* out of *6* test sessions in the MC vocabulary measurement. Thus, it cannot be concluded with certainty that dual glossing modes were completely more effective than the single glossing mode; however, the superiority of dual versus single mode is evident for the majority of test sessions in the two vocabulary tests.

The next section presents the discussion of glossing modes in short and long-term. This section is divided into two parts. In the first part, I relate the effectiveness of dual glossing mode over single glossing condition to the current survey in the literature. I then rationalize the efficacy of the single glossing mode in a few test sessions, followed by the reasons for the superiority of video/animation glossing format over audio glossing condition. Next, I justify some of the findings based on the possible explanations such as the consistency of instruction type and assessment method as well as a comparison between receptive versus productive vocabulary measurements. In the second part, I discuss the findings according to the two underlying theoretical frameworks of this study.

Part One: (a) Relating Findings to the Survey in Literature

Effectiveness of dual glossing modes over single glossing mode

With the advancement and incorporation of technology into the domain of vocabulary acquisition via glossing strategy, glosses have gone beyond the simple textual definitions (Chen, 2016; Chen & Yen, 2013) to multimedia glosses that include the addition of verbal (word definitions & audios) and visual (static pictures, graphics, videos, and animations) modes (Al-Seghayer, 2001; Chun & Plass, 1996; Sadeghi et al., 2016; Sakar & Erçetin, 2005; Salem, 2006; Türk & Erçetin, 2014; Yanguas, 2009), to name just a few. Glossing word items through the use of different annotation types can improve L2 acquisition "when presented digitally on a computer screen" (Al-Seghayer, 2016, p. 179). Empirical evidence to date revealed positive effects of different types of multimodal glossing on L2 learners' vocabulary knowledge (Sadeghi et al., 2016; Choi, 2016) and retention (Salem, 2006; Al-Seghayer, 2001). Also, Mohsen and Balakuma's (2011) meta-analysis revealed a positive and facilitative effect of multimedia glosses on vocabulary acquisition.

The results of the present study are in tune with several studies in regard to computer-based multimedia glosses, showing that L2 learners can acquire target words better when they are given a dual combination of gloss modes (e.g., L2 definition & pictures; L2 definition & audios; or L2 definition & video/animations) than when they consult definitions alone or with no gloss mode (Abraham 2008, 2007; Akbulut, 2007;

Al-Seghayer, 2001; Chun & Plass, 1996; Lin & Tseng, 2012; Nagata, 1999; Plass, et al., 1998; Sadeghi et al., 2016; Tabatabaei & Shams, 2011; Yeh & Wang, 2003; Yoshii, 2006; Yoshii & Flaitz, 2002; Yun, 2011). As an example, while Chun and Plass (1996) revealed that the combination of textual definition and pictures had a facilitative effect on participants' incidental vocabulary learning, Al-Seghayer (2001) and Lin and Tseng (2012) found that the dual combination of text[55] and video led to better incidental word learning and retention. Though the modes of presentation were differently effective in each study, the dual glossing mode was still particularly more effective than the single glossing mode. However, in Akbulut's (2007) study, there were no significant differences between the two visual modes of text and pictures and/or text and videos. Recent meta-analyses on the effect of hyper-text/computer-based glosses on vocabulary acquisition showed a moderate to large magnitude of the gloss combination of text and visuals for word learning (Yun, 2011; Abraham, 2008). Thus, the superiority of the dual glossing modes over the single mode in many vocabulary test sessions of this study lend support to the available literature suggesting that bimodal glossing is more effective than single mode glossing in aiding/promoting learners' vocabulary acquisition and reinforcing their retention (Al-Seghayer, 2001; Chun & Plass, 1996; Jones, 2003). However, as the findings showed, incorporating visual annotations to texts for word learning is not always effective (Kim & Kim, 2012); meaning that the single glossing mode (i.e., L2 definition alone) was also effective for some test sessions in both the PR and MC vocabulary measurements of this research. The next section discusses the rationale based on the following premises.

[55] Text refers to definition of the target word in L1 or L2.

Effectiveness of single glossing mode in few test sessions

First, similar to other studies, it is probable that the participants still considered the provision of simple textual glosses essential for the comprehension of the target words than the other different gloss types although they were exposed to different gloss modalities (Acha, 2009; Boers et al., 2017; Chen, 2016; Chen & Yen, 2013; Chun & Plass, 1996; Davis & Lyman-Hager, 1997; Erçetin, 2003; Farvardin & Biria, 2011; Ko, 2005; Laufer & Hill, 2000; Lomicka, 1998; Yoshii, 2014). This means that word definition was still the mode which resulted in a better vocabulary learning (Al-Seghayer, 2001; Chun & Plass, 1996; Kim & Gilman, 2008) and retention (Kim & Gilman, 2008) for the participants of the experimental groups in comparison to the control group. In Lomicka's (1998) pilot study, the efficacy and use of different glossing conditions (e.g., no glossing, L2 definition and L1 translations, images, references, questions, and pronunciation) was investigated for L2 learners' reading comprehension. The findings revealed that the intermediate-level participants preferred definitional glosses to other gloss types, most likely because they did not realize "the relevance of the other glosses in helping them with the reading process" (p. 48). Working with different group of participants and language abilities, Acha (2009) conducted an experiment on the third and fourth grade children in order to examine if simultaneous verbal and visual presentation modes (i.e., L1 definition, picture, or both) would affect learners' vocabulary acquisition in short and long-term. The results showed that the verbal annotation mode was more effective than the visual mode for word learning and recollection after two weeks. The rationale could be attributed to children's abilities to process texts with pictures or without illustrations. Participants with high verbal and visual abilities may benefit better than those with low abilities from concurrent multimedia presentation of word definitions and pictures (Mayer & Sims, 1994).

Second, it is possible that the visual annotations (L2 definition & video/animation glossing) of some words may not have been clear enough for the participants to understand the meaning of the word. As an example, an animation clip for the target word *feather* shows an object (i.e., feather) in motion. However, a feather moving in circle might represent the feather itself to one learner, but be interpreted as the act of spinning to another learner. Another reason can be justified based on the cognitive load theory (Chandler & Sweller, 1991), suggesting that adding visual representations to the verbal mode may generate a higher and more extraneous load than presenting verbal annotation alone (i.e., text-definition) (Mohsen & Balkumar, 2011; Acha, 2009). It is possible that the provision of L2 definition & video/animation glossing, as the visual representations, may have involved a higher cognitive load for the learners to process the text than when presenting with word definitions alone. Consequently, the limited working memory of the learners as well as the load of the information have resulted in participants' low performance in the sessions that involved dual glossing modes. Research has shown that the performance of adults with low cognitive ability was negatively influenced when visual mode of pictures was added to their multimedia learning/instruction (Acha, 2009; Boers et al., 2017; Plass et al., 2003). Therefore, the superiority of the single glossing mode in a few test sessions can be attributed to participants' different preferences for verbal or visual representations (Rassaei, 2017; Plass et al., 1998). In line with Plass and colleagues (1998), Kim and Kim (2012) indicated that some participants perform differently due to their high verbal ability and low visual ability to process the information in a text. Also, Rassaei (2017) found that learners' preferences to different learning styles affect their vocabulary learning.

Third, it is possible that participants' preferences for gloss modes depended on the target words to be glossed, and the type of text presented. In other words, the combination of L2 definition and audio glossing or L2 definition and video/animation glossing was not effective for all the target glossed words and all passages in the present study. As an example, the single glossing mode was significantly more effective than the other two glossing modes for text 3 in the PR test; but it was as effective as dual glossing mode for texts 2 and 3 in the MC vocabulary measurement. As for the variation, the finding might signify that the target words in text 3 were conveyed more effectively by means of textual definitions rather than the two other glossing modes. To further justify the claim, research has shown that text type/genre (e.g., narrative, expository, explanatory, authentic, modified) can play a role in affecting participants' performance in one mode rather than another (Farvardin & Biria, 2011; Ko, 2005; Lomicka, 1998).

Farvardin and Biria (2011) conducted a study, investigating the impact of different gloss types (e.g., L1, L2, and MC glosses) on reading comprehension across two text forms (i.e., expository & narrative). The findings showed that it was L1 gloss that facilitated learners' reading comprehension in the narrative text the most; and L2 and MC glosses that yielded the highest effect on participants' comprehension in the expository text. Yet, Ko (2005) found a reverse result showing the effectiveness of L2 glosses for narrative texts. However, researchers such as Jacobs and colleagues (1994), and Yoshii (2006) who used both L1 and L2 to investigate the type of annotation that best facilitated L2 vocabulary learning found that neither L1 nor L2 gained significant differences in favor of one text type over the other. Paribakht and Wesche (1999) found that the *scientific* text type that they chose for reading comprehension influenced the participants' performance in reading comprehension questions. They further stated that:

> Text characteristics that evidently influenced learners in terms of both their motivation and their success in guessing meanings included the topic, informational content, and genre of the reading text. (Paribakht & Wesche, 1999, p. 210)

Thus, depending on how the learners perceived the difficulty level of the text, or the chance of inferring the meaning of the target words based on some textual or contextual cues as well as participants' previous knowledge (Paribakht & Wesche, 1999), their choice of glossing mode(s) might have differed. However, learners' proficiency level (Lomicka, 1998) and type of assessment or outcome measures (Chen, 2016; Lomicka, 1998) also affect the selection of one glossing mode over another. In a recent meta-analysis on computer-mediated glosses, Abraham (2008) referred to three influential factors that influence the choice of gloss types: "level of instruction, text type, and task of assessment[56]" (p. 199). Regarding the text type, the analysis showed that narrative text gained a larger effect in comparison to the expository text with medium effect in multimedia setting. The present study used three expository reading passages[57]; however, as the analysis of text types was not within the scope of this study, the discussion relevant to this section, mentioned above, is based on the researcher's interpretation, and thus requires further research investigation.

Finally, another reason that might contribute to the efficacy of a single glossing mode over a dual glossing on some test sessions is the type of video/animation clips

[56] This variable is discussed in the subsequent section of this chapter.
[57] Review methodology chapter for the reading passages used in this study.

selected to instruct the glossed words. Although all the clips were piloted twice before the actual study, it is probable that they were not as transparent or clear as they should be for some of the target words in one text; thus, the participants preferred word definition mode (single glossing) over word definition and video/animation glossing (dual glossing). In other words, learners were not able to clearly understand the meaning of some of the glossed words presented via video/animation clips, and their scores were affected in the delayed vocabulary tests resulting in low performance. As an example, the findings showed that video/animation glossing mode was not effective for the participants in day 3 of the instruction. As a result, they performed low on the delayed post-tests (i.e., long-term effect) for both the PR and MC productive recognition tests. It might be probable that the video/animation clips that were provided for the target words in text 3[58] were not clear enough for the participants; consequently, they did not help the learners to retrieve the words in long term.

Nevertheless, although the single glossing mode showed to be more effective than dual glossing modes in a few test sessions, this does not deny the efficacy of combining word definitions with audio glossing or video/animation glossing for vocabulary learning and retention. As the sub-test comparisons showed, in both PR and MC productive recognition vocabulary tests, the dual glossing modes were also effective for L2 learners' vocabulary learning in short-term, and partially effective for their long-term retention. However, among the two dual glossing modes, as far as short-term retention was concerned, the dual modes of audio and video/animation glossing were effective in all instructional conditions for PR and MC tests; whereas video/animation glossing was

[58] See Appendix F for a sample of text 3.

effective in 4 out of 6 instructional conditions for both PR and MC tests in long-term (2 cases in PR test, and 2 cases in MC test)[59]. The next section justifies the reason(s).

Effectiveness of video/animation glossing over audio glossing mode

The findings of the present study showed that video/animation glossing mode was significantly more effective than audio glossing condition for several test sessions in both the PR and MC productive recognition tests; thus, supporting the studies that showed the effectiveness of visual representations for vocabulary learning and retention (Lin & Tseng, 2012; Al-Seghayer, 2001). To put it differently, audio glossing was effective for *1* out of *6* test sessions in the PR, and *1* out of *6* test sessions in the MC productive recognition tests, meaning that the bimodal glossing of L2 definition and audio recording was not as effective as the dual glossing of L2 definition and video/animation. The findings are in line with some studies on the inefficacy of audio inclusion/addition to present multimedia information (Sakar & Erçetin, 2005; Yeh & Wang, 2003).

One reason why the audio glossing condition was less effective could be related to the limited capacity of learners' working memory[60], which impedes the processing of audio input in mind; thus, "learners predominantly rely on visual input" (Al-Seghayer, 2016, p. 181) of video/animation than the verbal input of audio recording to understand the passage, and learn the new words.

Second, the relative ineffectiveness of audio glossing mode might be interpreted according to how Mayer (2014) defines *multimedia*. In his definition, multimedia refers to "presenting both words (such as spoken text or printed text) and pictures (such as

---

[59] See Tables 5.1 and 5.2 for the comparisons.

[60] One of the assumptions of cognitive theory of multimedia learning is suggesting that each channel has a limited capacity to process received information (See Chapter Two for full discussion).

illustrations, photos, animations, or video)" (Mayer, 2014, p. 2). Following the same explanation, audio recording or *spoken text* and word definitions or *printed text* go to the same category of word in "*verbal form*" (Mayer, 2014, p. 2). Thus, presenting the definition of a new word along with its audio recordings (two components of the verbal forms) at the same time might have induced an unnecessary distraction (Kim & Gilman, 2008) and a redundancy effect[61] (Sweller, 2010; Yeung, 1999) for the learners to process nonessential information in the text—leaving them with "a heavier cognitive load" (Kim & Kim, 2012, p. 63). Mayer and Fiorella (2014) refer to this condition as "a redundant multimedia presentation" (p. 287), since learners were exposed to the same target words in both spoken and printed forms.

The reason can be further clarified based on the cognitive theory of multimedia learning, which suggests that multimedia information may not be effective if one channel (i.e., verbal or visual) is overloaded (Sakar & Erçetin, 2005) with several information of the same category or type. The justification sounds plausible in the case of the present study where the verbal channel carries both the definition of the word as well as its audio recording, and both belonging to the category of verbal form. Learners then have to process the two-verbal information received from the text through the ears and eyes without receiving any visual input from the eyes (visual channel). Thus, the overloading of this single channel, and the provision of the additional input, might have created redundancy (Mayer, 2001), affecting participants' vocabulary learning and retention. However, with the video/animation glossing mode, learners' verbal and visual channels are both challenged and activated, since learners are exposed to verbal (text definition)

[61] The redundancy effect occurs "when unnecessary, additional information is presented to learners" (Sweller, 2010, p.130).

and visual (video/animation) information at the same time, allowing them to receive the input verbally from the ears and eyes, and visually from the eyes (Mayer, 2014, 2007). Therefore, presenting learners with several elements to be processed in visual or verbal working memory can lead to cognitive overload (Mohsen & Balkumar, 2011; Chandler & Sweller, 1991) or redundant information (Mayer & Fiorella, 2014). However, to minimize the effect of high cognitive load in case of video/animation mode, the participants received the information coming from two different channels (verbal & visual) simultaneously (Mayer & Fiorella, 2014; Türk & Erçetin, 2014).

Another discussion that deserves attention in the present research is the efficacy of the video/animation mode in comparison to the audio glossing format. Although the two annotation types were subject to redundancy and extraneous cognitive load, one might ponder if presenting both verbal and visual information (e.g., L2 definition & video/animation) together at the same time have again created redundancy and distraction effects for the participants. The discussion could rely on the findings of some studies showing that video/animation clips are effective learning instruments that foster word learning due to their rich contextual clues (Al-Seghayer, 2001), leaving learners with "a more memorable experience and, in the long run, a better retrieval cue (Al-Seghayer, 2001, p. 224). Additionally, presenting target vocabulary via relevant video/animation clips triggers participants' curiosity (Al-Seghayer, 2001) to follow what comes next and builds mental images of the target words by constructing form-meaning relations (Mayer, 2014); thus, enhancing the vocabulary learning (Al-Seghayer, 2001) and retention (Lin & Tsneg, 2012).

In sum, the dual mode of L2 definition and video/animation glossing was more effective than bimodal annotation of L2 definition and audio glossing. This supports Underwood's (1989) commonly-cited statement that "we remember images better than words; hence we remember words better if they are strongly associated with images" (p. 19).

Consistency of type of instruction and type of assessment

The findings of the present study showed that, in general, dual glossing modes were effective for L2 learners' word learning and retention in both PR and MC productive recognition tests; however, it cannot be stated with certainty that the dual modes were more effective than single condition, because the L2 definition alone was also influential in L2 learners' word learning and retention in some test sessions.

One factor that might have affected the findings of the present study in this regard is the inconsistency of the mode of vocabulary presentation with the method of assessment. The participants were instructed via the multimedia glossing formats in a computer-based (non-traditional) learning environment where the glossed words were presented on a large screen with both verbal (L2 definition & audio) and visual (L2 definition & video/animation) elements available. However, paper and pencil vocabulary tests (traditional format) were given to them to measure their vocabulary knowledge. If the teaching and testing modes were consistent, the learners' test performance might be different (Akbulut, 2007; Ariew & Erçetin, 2004). Akbulut (2007) indicates that if learners are presented with non-traditional vocabulary instruction, but are evaluated with traditional measurement tools, "the results might be affected" (p. 515). Studies have also shown that learners' performances have changed for the best if the instructional mode of lexical items matched the method of vocabulary measurement (Chun & Plass, 1996; Jones, 2004; Jones & Plass, 2002). Chun and Plass (1996a) investigated the impact of multimedia glossing annotations on L2 vocabulary learning from two written production and recognition measurements, using both pictorial and verbal representations, which matched the modality of word presentation. They observed that when the testing mode paralleled learners' vocabulary acquisition, their performance noticeably improved resulting in a higher rate of correct answers (i.e., 77%) for both immediate and delayed vocabulary tests. Similarly, conducting two consecutive studies, Jones (2004) investigated the effectiveness of verbal (written/translation text) and visual (pictorial text) multimedia annotations in an aural setting on learners' acquisition of L2 (French) words. In her first study, the participants were required to listen to a text with the possibility of clicking on the annotated words for written definitions/translations in verbal group, pictorial annotation in visual group, and both annotations in written and pictorial group. The learners were then measured via immediate and delayed recognition vocabulary tests. The findings showed that the participants of the experimental groups performed better in recognizing English words in written, visual, or both formats than the control group regardless of the testing method, meaning that participants recognized the words equally in both pictorial and written representations in recognition tests. However, in the second study, the same procedure for the vocabulary instruction followed with a change in the mode of assessment, meaning a recall vocabulary test was substituted for the recognition test as the immediate and delayed post-tests. Interestingly, the findings revealed that the group in written annotation mode did very well on the recall test and produced English translations of the French words due to the consistency of vocabulary teaching mode and assessment method. In other words, the participants with access to written annotations or both outperformed those without annotations in production test. Thus, matching the testing mode with instructional mode affected the test results and assisted the students to learn more vocabulary (Jones, 2004). Additionally, Al-Seghayer (2016) notes that the uniformity between vocabulary instruction and evaluation establishes a link between how target glossed words are presented to the participants and how they are tested. In a recent meta-analysis on computer-mediated glossing, Abraham (2008) referred to the type of assessment as an "important moderator of outcomes" on learners' incidental vocabulary learning (p. 210). Thus, incorporating both verbal and visual elements of the vocabulary instruction into measurement methods would be more convenient and useful in multimedia learning environment, leading to authentic results (Akbulut, 2007).

Comparing productive recall versus multiple-choice productive recognition vocabulary measurements

According to the findings of the present research, the participants in the experimental groups increased their scores from pre- to immediate post-tests in both PR and MC productive recognition tests; but also, they revealed a significant decrease in their scores from the immediate to delayed post-tests two weeks later. However, learners still had some gains in scores from the pre- to delayed post-tests in vocabulary measurement tools. The results are in line with some studies suggesting that the participants did not show word gain from immediate to delayed post-test (Arpaci, 2016; Yanguas, 2009) in production and recognition vocabulary tests, but revealed significant improvements from pre-to delayed tests (Yanguas, 2009). Additionally, according to the descriptive tables[62] shown in chapter *four*, it was revealed that the word gain was higher from the pre-to delayed post-tests in the MC productive recognition test compared to the PR test[63]. The rationale could be justified based on the types of measurements used to gauge learners' vocabulary knowledge. The MC productive recognition test in this study was in the form of active recognition (Laufer & Goldstein, 2004) in that the participants were given the meaning of the target words in L2, and they were required to choose the correct word among the four options. Also, the recall test was in the form of asking the learners to retrieve the words from their memory (Laufer & Goldstein, 2004), and produce them in the space provided.

Research has shown that MC test formats, as the most appropriate type of measurement, are commonly used to assess participants' receptive/passive word knowledge[64] (Stewart, 2012), and usually strengthen participants' existing memory (Jones, 2004). On the other hand, recall tests, as a form of measuring productive/active word knowledge, are more demanding since "the learners must search for the correct response within their mental representations of the newly experienced information" (Jones, 2004, p. 124). However, despite their difficulty level, productive tests are an accurate means of diagnosing one's language proficiency (Stewart, 2012; Laufer & Goldstein, 2004). The reason MC tests are more popular than the production tests are their "practicality and cost-effectiveness" (Stewart, 2012, p. 54), quick administration and preparation (Öztürk, 2007), ease of scoring (Gyllstad, Vilkaite, & Schmitt, 2015;

---

[62] See Tables 4.1 & 4.11 (in sum & sub-test comparison for PR), & 4.22 & 4.32 (in sum & sub-test comparison for MC) in chapter *four*.

[63] This was achieved through tracing the mean scores of PR and MC productive recognition test from pre-tests to the delayed post-tests in each group.

[64]Passive knowledge refers "the ability to supply the word meaning" and active knowledge refers to "the ability to supply the word form" (Laufer & Goldstein, 2004, p. 406)

McAllister & Guidice, 2012; Öztürk, 2007), and measuring a large number of words in short time (Gyllstad et al., 2015). Yet, they pose several challenges to vocabulary assessment. The very first challenge is the probability of guessing in MC tests (Gyllstad et al., 2015) that affects the reliability of the measurement. Webb (2005) remarks that any 4-option multiple-choice format on the receptive tests are subject to "a 25% chance" (p. 49) of guessing, to which Wesche and Paribakht (1996) agree. The next criticism is that MC vocabulary tests do not assess learners' knowledge of lexical items in depth (Schmitt, 1999); rather they evaluate "knowledge of the distractors" (Gyllstad et al., 2015, p. 278). In other words, learners' word knowledge, including both form and meaning (Nation, 2001), are not sufficiently measured via MC vocabulary tests. Interestingly, McAllister and Guidice (2012) state that "multiple choice tests are often tests of literacy rather than a meaningful assessment of learning" (p. 194).

One possibility for the higher scores of the participants of the present study in the MC test from those in the PR test (i.e., from pre- to delayed post-test) is related to the 25% chance for the learners to choose the item(s) by correct prediction/guessing (Meara & Buxton, 1987; Webb, 2005; Wesche & Paribakht, 1996).

Furthermore, taking the idea from Webb (2005), it is likely that no significant differences were found on MC productive recognition tests, in the present study, as a result of the ceiling effect, because the MC test might have been easier for the learners to deal with than the PR test—thus, affecting the test scores. Since multiple-choice tests are easier to recognize and respond to (McAllister & Guidice, 2012) than productive vocabulary tools, it is probable that the participants could have answered MC test items more easily compared to the PR test, leaving no discriminatory effect between the two measurements. To put it differently, the MC vocabulary test, in this study, was not a strong measurement tool to distinguish participants' word knowledge in comparison to the PR test due to its easiness.

Another interpretation for the different performance of the participants in both PR and MC productive recognition tests is the allocated time to complete the task. Following the findings of the pilot study, learners were given 15-20 minutes to finish both vocabulary measurements; however, due to the different nature of the tests, it might have been possible for the participants of the actual study to complete the PR test longer than expected as they were struggling with retrieving the orthography of the target words in order to write the correct letters in the provided blanks. Yet, the same participants might have finished the MC vocabulary test at the right timing (Webb, 2005) or even sooner. Considering the productive and receptive vocabulary tasks as "a function of receptive and productive learning" (Webb, 2005, p. 49), research has shown that the productive task takes longer than receptive one to be completed (Waring, 1997b). Thus, the low performance of the learners in PR test might be attributed to the lack of time they had to accomplish the task. Reviewing the notes regarding the actual data collection stage, I also witnessed that if participants had more time to do the PR test, their performance might have been positively affected.

Also, regarding the glossing modes, the results demonstrated that the three glossing annotations were differently effective in PR and MC productive recognition tests for both immediate and delayed post-tests. As an example, in the MC vocabulary measurement, all three glossing modes were similarly effective for learners' short-term retention; whereas the trend of effectiveness was different on each instructional day for the PR test (dual glossing modes on days 1 & 2, and all glossing modes on day 3). Again, this fluctuation could be attributed to the type of measurement used in this study. Learners seemed to have done better on the MC productive recognition tests than PR tests as far as immediate tests were concerned. The findings are in tune with Abraham (2008) who found “a comparatively larger effect for receptive than for productive tests that was sustained over time” (p. 211), suggesting that learners’ receptive or passive word knowledge is greater than their active or productive vocabulary (Laufer & Paribakht, 1998; Waring, 1997a; Webb, 2005). Another possible explanation could be that with productive recall tests, learners have to rely on their memory to elicit the words (Turk & Ercetin, 2014), thus affecting the test results; however, MC productive recognition tests are easier to process and are “more readily gained” (Yusuf et al., 2014, p. 106).

In general, PR tests seemed to be more challenging than MC productive recognition tests for the participants of the present study as they had to struggle with retrieving the target words from memory and writing them with correct orthography in rather a short time. Considering the 25% chance of guessing in MC vocabulary tests, it sounds reasonable that the participants scored lower in the PR test from pre-to delayed post-tests compared to their performance on MC productive recognition tests. Moreover, it is probable that due to ceiling effect, MC test results were not discriminatory enough to assess learners’ accurate amount of word learning and retention.

The discussion thus far has related the findings of the present study to the current literature. The next section discusses the efficacy of dual versus single glossing modes in light of the theoretical frameworks of this study.

Part Two: (b) Relating Findings to the Theoretical Frameworks

Effectiveness of dual glossing modes over single glossing mode

The results of this study implied that the availability of visual and verbal annotations, along with textual definitions, assisted L2 learners to perform better on vocabulary tests than a single annotation type alone for the majority of test sessions. The two underlying theoretical frameworks of the present study, dual-coding theory (Paivio, 1986) and cognitive theory of multimedia learning (Mayer, 2014, 2007), supported the efficacy of dual glossing modes versus single glossing mode. According to Paivio (1986), the two independent, but interrelated verbal (text, spoken narratives/audio recordings) and visual (pictures/illustrations/videos & animations) channels help process the information dually through making referential connections between them. The referential connection of the two modes has an additive impact on learners' recall (Mayer & Anderson, 1991) and "complement each other in facilitating retention of information" (Akbulut, 2007, p. 500). Paivio (1986) remarks that learning occurs effectively and better if both verbal and visual information are dually presented to the learners rather than only one; learners are then able to construct referential connections between the two forms for a successful word learning (Al-Seghayer, 2001). Likewise, Mayer (2001) states that "presenting an explanation with words and pictures[65] results in better learning than does presenting words alone" (p. 78). The rationale for the better performance of the participants in the dual glossing conditions is also attributed to the fact that when learners encoded information (i.e., target words) in both visual and verbal formats, they had an opportunity to process the words in two channels; and thus, retrieve them better in

[65] See the Introduction chapter for Mayer's definition of picture.

more than one way. Oxford and Crookball (1990) explain that when a reading passage is annotated with pictures, it can access more parts of the brain leading to a deeper word processing than when one mode is presented. In line with dual-coding theory, Mayer (2014) indicates that for meaningful word learning to occur in a text in multimedia learning setting, learners must first select relevant words and images from the input (i.e., text) they receive in both verbal (written or auditory) and visual (pictorial) channels; and then organize the information into coherent verbal and visual representations in order to send it to the working memory; finally, learners should integrate the verbal and visual information into their prior knowledge. However, for the information to stay in long-term memory, they have to actively move back and forth from long-term memory to working memory, building referential connections between the two formats (Mayer, 2014, 2001, 1997; Jones, 2004).

From the theoretical perspective, relating the findings of the present study to Mayer's cognitive theory of multimedia learning and Paivio's dual-coding theory, it can be implied that the presence of the two separate, but interrelated, verbal (L2 definition & audio) and visual (L2 definition & video/animation) information assisted the participants to establish direct mental connections between the two models in short-term memory, and paved the way for effective long-term word retrieval. In other words, participants benefited more when they received the target words through the verbal tools of L2 definition and audio as well as the visual tools of L2 definition and video/animation than L2 definition alone in a multimedia-based learning setting. Additionally, the two notions of *transfer* and *retention* in a multimedia learning environment (Mayer, 2001) may contribute to the better performance of dual mode conditions over a single mode setting.

Transfer refers to learners' ability to utilize the information to solve new problems; and retention happens when the information is retrieved in the multimedia presentation (Mayer, 2001). In the current study, transfer was achieved through participants' attempt to do PR and MC productive recognition tests; and retention was obtained via using delayed post-tests. Thus, the results showed that combining text and audio or text and video/animation led to a tentatively better performance of the participants in PR (*4* out of *6* sessions) and MC (*3* out of *6* sessions) tests in both immediate and delayed post-tests (i.e., short & long-term retention); thus, the words that were dually coded were learned better than words with text definition alone. The findings are in line with studies that showed "combining definitions of words with associated visuals regardless of the type of visual used is more effective in facilitating vocabulary learning than providing only definitions of words" (Akbulut, 2007, p. 513). Therefore, the results support both Paivio and Mayer's theories suggesting that presenting an explanation of a word along with verbal and visual cues is better than solely in words. However, since the three glossing modes were shown to be equally effective in some test sessions (i.e., one session in PR & one session in MC tests), the justifications cannot be generalized to a large extent.

In favor of the efficacy of bimodal glossing over single mode, the two principles of multimedia learning, namely multimedia principle and temporal contiguity principle, contribute greatly to L2 learning (Türk & Erçetin, 2014), explaining why the participants who received target glossed words via simultaneous bimodal glossing performed better than those who were instructed via single glossing mode in the majority of test sessions. The next section discusses the principles in regard to the findings of this study.

Simultaneous display of multimedia glossing. The multimedia principle suggests that learners learn better when they are exposed to both verbal and visual information rather than either alone (Mayer, 2014, 2007). Conversely, the contiguity principle implies that learning materials can be presented to the learners successively or simultaneously (Mayer, 2005). The findings of this study might suggest that the participants showed better performance when verbal (L2 definition & audio) and visual (L2 definition & video/animation) information were presented simultaneously, reinforcing their learning. Therefore, it can be argued that simultaneous display of glossed words was an efficient mode that allowed learners to engage with the text and "utilize text resources to a great extent" (Türk & Erçetin, 2014, p. 15). Also, the presentation of definitions and audio or definitions and video/animation simultaneously indicates that the two modes allow the brain to process verbal and visual information in working memory, and retrieve them faster in long-term when required (Sweller, 2005). In other words, the participants were able to perform better on vocabulary tests when presented with the information at the same time. Another rationale to justify the presentation of gloss information at the same time in this study was to avoid overloading learners' cognitive capacity. Mayer and Fiorella (2014) suggested that one possible way to eliminate the need to hold the information in working memory for a long time is to present the multimedia materials at the same time. This way, the definitions and audio recording or the definitions and video/animation clips that were shown together were more likely to contribute to L2 learners' word learning, reduce loads of materials, and finally recall better than when L2 definition is presented first and audio recording later (or L2 definition first & video/animation next) (Mayer & Sims, 1994). Furthermore, as one of the pre-requisites of meaningful learning is to make connections between words and images, presenting the materials at the same time facilitates this link, and enhances long-term recollection (Mayer, 2008).

Section Three: Learners' Perceptions and Attitudes

Questionnaire & Semi-Structured Interviews

The findings of the questionnaire and face-to-face semi-structured interviews unanimously showed that the dual glossing modes (i.e., L2 definition & audio glossing, and L2 definition & video/animation glossing) were perceived as being more favorable than the single glossing mode (i.e., L2 definition alone) for L2 learners' word learning and retention in short and long-term. Ninety-four percent of the participants believed that the bimodal glossing of L2 definition and video/animation helped them to learn the words easier. Around Eighty-five percent stated that the incorporation of video/animation glossing with L2 definition aided them to recall the new words better in long-term, and finally, sixty-five percent of the learners were enthusiastic to use this glossing technique in future for their vocabulary learning.

In regards to the efficacy of dual mode of L2 definition and audio glossing, 76% of the participants rated this technique as a practical strategy for learning the new vocabulary easily; around 64% of them stated that the glossing mode of L2 definition and audio pronunciation helped them to remember the words easier than other glossing modes in long-term; and about 64% of the participants were interested to utilize it for their future word learning experience.

With respect to the single glossing mode (i.e., L2 definition alone), it was found that less than 10% of the learners thought this glossing mode assisted them to acquire the words easily; only 31% of the learners approved the use of this strategy for long-term use; and around 13% of the learners strongly wanted to practice with the L2 definition alone mode for their future vocabulary learning efforts.

As implied, the participants seemed to prefer video/animation glossing more than the two other glossing annotations, because the majority of them rated this annotation type as "*strongly agree*" or "*extremely helpful.*" The findings are consistent with studies that employed questionnaires and interviews to seek learners' perceptions and viewpoints towards the effectiveness and usefulness of annotation types (e.g., multimedia glossing) for word learning (Al-Seghayer, 2001; Sakar & Erçetin, 2005; Yeh & Wang, 2003). The 94% figure in the present study is close to Al-Seghayer's (2001) study who found that 86.6% of the participants rated video clip condition as the most helpful mode of vocabulary glossing.

The findings of the questionnaire regarding the audio glossing mode also showed that learners' visual inclination was stronger than their verbal preference (e.g., written or spoken text) for word learning (Yeh & Wang, 2003). As an example, in the present study, only 14 participants (16.86 %) *strongly agreed* to employ this glossing technique for their future vocabulary learning use. However, the learners had a positive perception towards the audio glossing mode accompanied with L2 definition and viewed it as their second favorable annotation type.

Textual definition alone was also not a widely favorable means of vocabulary instruction for the participants in comparison to the other two glossing modes of L2 definition and audio glossing and L2 definition and video/animation glossing. Less than 10% of the learners *agreed* that learning new words would be facilitated if accompanied by L2 definition alone. This finding matches other studies, suggesting that the single glossing mode was helpful for, at most, 10% of the participants (Al-Seghayer, 2001); and participants learned better with pictures than with word definition alone (Yeh & Wang, 2003).

One point that deserves special attention here is the participants' learning preferences to verbal and visual modes[66] of gloss presentation. According to the findings of the questionnaire, the dual mode of L2 definition and video/animation glossing was the most desirable word learning experience for 75% of the participants; whereas, the other 25% of learners favored either bimodal glossing of L2 definition and audio glossing or L2 definition alone. Although the majority of the learners adhered to the combination of textual definition and visual representations as their preferred mode (i.e., visualizers), for effective vocabulary learning to occur, it is imperative to take into account the learning preferences and styles (Rassaei, 2017; Plass et al., 1998) of all the learners in a class. This is because learning preferences can facilitate students' interaction with the teaching/learning material and the environment, enabling them to "extract information from it" (Plass et al., 1998, p. 27); thus, leaving the learners with an enjoyable learning experience. Likewise, Rassaei (2017) remarks that students with visual learning styles take advantage of visual or textual glosses whereas auditory style learners benefit from audio/spoken forms of glosses. Attending to a learners' individual learning preference would also clarify the question of "for whom is multimedia instruction effective?" (Plass et al., 1998, p. 25), and matching the learners' learning style to the method of instruction enhances learning (Rassaei, 2017). From the findings of question 13 in the questionnaire,

[66] Verbal gloss presentation entails L2 definition alone as well as L2 definition & audio glossing; and visual gloss presentation includes L2 definition and video/animation glossing.

it revealed that 25% of the participants belonged to verbalizer groups where they favored the verbal mode of vocabulary presentations, such as L2 definition and L2 definition and audio glossing, whereas, the rest of the participants (75%) were the visualizers in that they benefited more from the visual display of the materials and learning activities. Plass and colleagues (1998) remarked that verbalizer-visualizer dichotomy is one "dimension of learning preference" (p. 27). Besides, back to the available literature on this domain, textual information is regarded as verbal information, and visual information such as images, animations, and video/clips are considered as visual information (Plass et al., 1998).

In regards to the face-to-face semi-structured interviews, it seems that the dual mode of video/animation glossing was influential for the learners of the present study. The comments of some of the participants implied that video animation helped them to learn and retrieve the target words better after two weeks, because the visual scenes helped them to keep the words in their mind for later use. Thus, the participants perceived the video/animation mode, accompanied with textual definition, as an interesting, motivating, and important practice for the comprehension of the texts (Sakar & Erçetin, 2005; Al-Seghayer, 2001), as well as being a very enjoyable word learning experience (Erçetin, 2003).

In sum, it can be concluded that the three glossing modes were effective for L2 learners' word learning and retention; however, the two simultaneous glossing modes of L2 definition and video/animation glossing, and L2 definition and audio glossing were considered as the first and second preferred vocabulary presentation modes respectively for L2 learners' vocabulary learning and long-term recollection; and the single glossing mode of L2 definition alone was participants' last preferred choice. Furthermore, the findings imply that the presence of these sources of information (L2 definition & video/animation, and L2 definition & audio glossing) can support L2 learners' ability to establish a connection between visual and verbal models in short-term memory, and paves the way for the effective retrieval of words stored in the long-term memory. Therefore, having two separate but interrelated verbal and visual systems allowed the participants of this research to benefit more if they received the target words through the verbal tools of L2 definition and audio glossing as well as visual tools of L2 definition and video/animation in a multimedia-based learning environment.

Chapter Summary

In this chapter, I discussed the major findings of this research in light of the relevant literature and theoretical frameworks. First, the findings regarding the impact of glossing in short and long-term (i.e., in-sum test comparisons) were overviewed and discussed. I discussed that glossing was significantly effective for word learning and retention because it provided learners with adequate context, drew their attention to unfamiliar words, helped them process the text for comprehension, and saved learners' time and effort. Also, I explained that glossing could be effective for long-term retention only if adequate exposure was provided. Nagata (1999) remarked that "a one-day lesson is not sufficient to establish long-term retention" (p. 476). This discussion addressed research question 1 and its sub-questions.

Second, the results of the impact of glossing modes in short and long-term (i.e., sub-test comparison) were briefed and discussed. The rationale for the efficacy of dual glossing modes over single glossing format in several test sessions was discussed; followed by the reasons regarding the effectiveness of single glossing mode in few test sessions. Also, it was discussed why video/animation glossing mode seemed to be more effective than audio glossing condition overall. The justifications such as the uniformity of instructional type and assessment methods as well as the comparison between receptive and productive vocabulary measurements shed light on some of the findings. I addressed research question 2 and its sub-questions through this discussion.

The third section discussed the research findings of questionnaire and semi-structured interviews related to the third research question looking at the perceptions and attitudes of the learners regarding the use of multimedia glossing for word learning and retention.

In the next chapter, I conclude this dissertation by providing a summary of the major findings as well as a brief discussion on each section. I discuss the pedagogical and research implications, and make recommendations for future research in this field. The limitations that shadow this study along with my concluding remarks are also explained.

CHAPTER SIX: CONCLUSION

Introduction

The purpose of this research was to investigate the effectiveness of simultaneous multimedia glossing on L2 learners' vocabulary learning in terms of both short and long-term word retention, drawing on Paivio's (1986) dual-coding theory and Mayer's (2014, 2007) cognitive theory of multimedia learning. To this end, 132 adult intermediate language learners were assigned to one control (A) and three experimental groups (B, C, & D). The experimental groups received three reading passages in three different glossing modes of L2 definition alone, L2 definition and audio glossing, and L2 definition and video/animation glossing. The control group received the same texts with no glossing mode (i.e., no specific vocabulary instruction). A Vocabulary Levels Test (VLT), as well as two vocabulary pre-tests of productive recall (PR) and multiple-choice (MC) productive recognition tests were administered at the beginning of the study to balance out the homogeneity of the learners and gauge their familiarity with the target words respectively. The participants received the instruction in three sessions, every other day, during one week, and were then given two immediate vocabulary post-tests after each session to investigate their short-term word learning and retention. Two weeks later, the same post-tests with different ordering formats were utilized to measure L2 learners' long-term word recollection.

In the preceding chapters, I explained the underlying theoretical framework, relevant literature review, methodology, and study findings as well as a discussion of the results. In this chapter, I re-state each of the research questions and sub-questions with a brief overview of the major findings. This chapter is divided into the following five main sections: (a) summary of the major findings including both in-sum and sub-test comparisons as well as the questionnaire and face-to-face semi-structured interviews, (b) pedagogical implications of the study, (c) research implications with direction to future research, (d) study limitations, and (e) the concluding remarks, discussing how the present research filled the gap in the current literature.

Summary of the Major Findings

Two main research questions, each with two sub-questions, guided this study. The first question asked about the effectiveness of different glossing modes on L2 learners' vocabulary learning over short and long-term word retention (in-sum comparisons). The second research question asked about which glossing mode(s) was effective for L2 learners' short and long-term word learning and recollection (sub-test comparisons). The next section reviews the major findings of each question with respect to the in-sum and sub-test comparisons, bringing supports from the survey in the current literature and theoretical frameworks respectively.

Support from the Survey in Literature

In-Sum Test Comparison

To address the first research question, two sub-questions were created. The first sub-question asked if glossing was effective for L2 learners' short-term word retention; and the second sub-question investigated the impact of glossing on participants' long-term word recollection. In order to achieve this goal, participants' sum of the scores were compared, once from the pre-to immediate post-tests for short-term word retention, and another time from the pre-to delayed post-tests for long-term word retention (i.e., in-sum test comparison including both between and within participant comparisons). The findings of sub-questions for both PR and MC productive recognition tests showed that the vocabulary technique of glossing was significantly more effective than non-glossing strategy for participants' short-term word learning and recollection (i.e., sub-question 1.1); but was partially effective for learners' vocabulary acquisition in long-term (i.e., sub-question 2.1). The findings are in line with previous research showing that glossing is a practical vocabulary strategy to enhance word learning and retention in short-term (Al-Seghayer, 2001; Chun & Plass, 1996); and facilitates vocabulary acquisition in long-term if words are adequately reinforced after being learnt (Jacobs et al., 1994; So, 2010; Yoshii, 2014) through repetition and exposure to the words (Arpaci, 2016) as well as doing several post-reading and word-focused activities/exercises after the instruction (Newton, 2013; Paribakht & Wesche, 1996; Webb & Chang, 2015).

Sub-test Comparison

In order to respond to the second research question, and similar to the first question, two sub-questions were created, and the impact of different glossing modes were investigated once by comparing the participants' sub-test scores (i.e., between-participant comparison) from the pre-to immediate sub-tests for short-term effect (i.e., sub-question 2.1), and another time from pre-to delayed sub-tests for long-term impact (i.e., sub-question 2.2). The findings revealed that, as far as PR vocabulary measurement was concerned, the dual glossing modes (i.e., L2 definition & audio glossing, and L2 definition & video/animation glossing) were effective for *two* out of *three* test sessions in short-term learning and word retention (days 1 & 2); and all three glossing modes were equally effective for *one* out of *three* test sessions (day 3) in the research. Also, the glossing modes were similarly effective for participants' short-term word learning and retention in MC productive recognition tests. The findings addressed sub-question 2.1. However, in terms of long-term word retention, the results of PR test showed that the dual glossing mode of L2 definition and video/animation was effective in *two* out of *three* test sessions (days 1 & 2), whereas the single glossing mode (i.e., L2 definition alone) was only influential for *one* out of *three* test sessions (day 3). In contrast, the dual glossing modes were effective for all *three* test sessions of MC productive recognition test in terms of L2 learners' word retention in long-term with single glossing mode being positively influential for *two* out of *three* test sessions (days 2 & 3). Although the superiority of dual versus single glossing modes is evident for the majority of test sessions in the two vocabulary tests, the findings cannot be generalized with certainty that dual glossing modes were completely effective since single glossing mode was also shown to be influential for learners' word learning and retention in both short and long-term in few test sessions.

Nevertheless, the findings of both short and long-term for PR and MC productive recognition tests are aligned with other studies showing that the simultaneous dual presentation modes are more effective than single glossing mode in helping learners with vocabulary learning and retention (Al-Seghayer, 2016, 2000; Chun & Plass, 1996; Lin & Tseng, 2012; Sadeghi et al., 2016; Salem, 2006; Yeh & Wang, 2003; Yoshii & Flaitza, 2002). On the other hand, the outperformance of the participants receiving L2 definition alone in a few test sessions also confirms the findings of some studies regarding the effectiveness of textual word definition for vocabulary acquisition and recollection (Acha, 2009; Boers et al., 2017; Chen & Yen, 2013; Chen, 2016; Ercetin, 2003; Farvardin & Biria, 2011; Kim & Gilman, 2008; Ko, 2005; Yoshii, 2014), to name just a few. In general, the following factors could be listed regarding the efficacy of single glossing mode over dual glossing in some few test sessions: **(a)** participants' lack of attention to the visual annotations, **(b)** a higher cognitive load for the learners to process the text when presented with dual glossing modes than with word definitions alone, **(c)** types of target words to be glossed, and type of text to present, **(d)** learners' proficiency level, **(e)** type of assessment/outcome measures, and **(e)** type of video/animation clips selected to instruct the glossed words.

The findings also revealed that among the dual glossing modes, the simultaneous presentation of L2 definition and video/animation glossing led to a better word learning and retention than L2 definition and audio glossing. The rationale could be attributed to the factors such as: (a) limited capacity of learners' working memory that blocks the process of audio input in mind; (b) unnecessary distraction and a redundancy effect for the learners to process nonessential information in the text, which might have left them with a heavy cognitive load; and (c) rich contextual clues of the video/animation clips that triggered learners' curiosity, building comprehensive form-meaning relations.

Another factor that might have affected the findings of the present study in regards to the efficacy of the dual versus single glossing modes is the inconsistency/non-uniformity of the display mode of vocabulary presentation (i.e., non-traditional vocabulary instruction by means of multimedia glossing) with the method of vocabulary assessment/measurement (i.e., traditional way of testing such as paper and pencil). Research has shown that if the two modes are not consistent, participants' performance might be affected (Akbulut, 2007; Jones, 2004; Jones & Plass, 2002).

Another subject that was fully discussed in the previous chapter[67] was the changes/fluctuations in the participants' scores from the pre-to immediate, immediate to delayed, and pre-to delayed vocabulary post-tests. The findings revealed that learners had some improvements from pre-to immediate tests, showed some loss in scores from immediate to delayed, and had some word gains from pre-to delayed post-tests in vocabulary measurement tools. Besides, the word gain from pre-to delayed post-tests was higher in MC productive recognition test compared to the PR test.

The possible reasons could be traced to: **(a)** the type of testing tools that were used to assess learners' vocabulary learning after the instruction and two weeks later. MC vocabulary tests, as a means of evaluating participants' receptive/passive word knowledge (Stewart, 2012), are subject to the 25% chance for the participants to select the item by correct guessing (Webb, 2005), which resulted in learners' higher scores on the delayed post-tests in comparison to the PR vocabulary test; and **(b)** as a result of the ceiling effect, it was likely that MC productive recognition test might have been easier for the learners to complete than the PR test, and could not adequately measure participants' accurate amount of word learning and retention; Also, **(c)** similar timing spent on each vocabulary task (15-20 minutes) was also another factor that could have intervened to the higher scores of the participants on MC productive recognition delayed post-test in comparison to the PR test. Although the results of the pilot tests showed that

[67] See discussion chapter for more details.

the estimated timing to complete each test was suitable, it might have been possible for the participants of the actual study to need more time to finish the PR test than given as they were dealing with recalling the spelling of the target words in order to write the correct letters in the provided blanks; thus 15-20 minutes were not sufficient for them to accomplish the PR task, resulting in lower scores compared to MC productive recognition measurement; and finally, **(d)** the reason that the three glossing modes were differently influential in PR and MC productive recognition tests for both immediate and delayed post-tests might be contributed to the nature of the tests. The PR test required the participants to rely on their memory to elicit the words (Türk & Erçetin, 2014), whereas MC productive recognition tests were easier for the learners to process because they were "more readily gained" (Yusuf et al., 2014, p. 106).

Support from Theoretical Frameworks

As the findings showed, generally, bimodal glossing was more effective than single glossing mode for L2 learners' word learning and retention for the majority of test sessions. Dual-coding theory (Paivio, 1986) and cognitive theory of multimedia learning (Mayer, 2014, 2007), the two underlying theoretical frameworks of the present study, lent support to the effectiveness of dual glossing modes versus single glossing mode, meaning the two verbal and visual representation modes resulted in a more effective learning experience than only one. The reason was based on the fact that when participants received the target words in two formats (i.e., visual and verbal), they were able to process the words in two channels, resulting in better long-term retrieval. In other words, the two separate, but interrelated, verbal (L2 definition & audio) and visual (L2 definition & video/animation) representations aided the participants to build direct mental connections between the two models in short-term memory, and paved the way for effective long-term word retention. However, as the findings revealed, the three glossing modes were equally effective in some test sessions (i.e., one session in PR & one session in MC tests); thus, the justifications cannot be generalized to a large extent/largely. Additionally, referring to the two multimedia learning principles, the multimedia and temporal contiguity principles[68], it was implied that the participants who received the instruction through simultaneous dual glossing modes performed better than those who were taught via single glossing mode in the majority of test sessions. Multimedia principle suggests that learners learn better if both words (written and spoken or either) and pictures (images, icons, graphics, animations and video), as two sources of verbal and visual representations, are presented to them (Mayer, 2014, 2007); and temporal contiguity principle denotes that learning can be facilitated if learners are exposed to the learning materials successively or simultaneously (Mayer, 2005). Besides, displaying the information simultaneously would eliminate the possibility of overloading learners' cognitive capacity, which is limited.

The next section summarizes the findings of the questionnaire and face-to-face semi-structured interviews, relying on L2 learners' perceptions and attitudes about the efficacy of simultaneous multimedia glossing on word learning and retention.

[68] Review chapter one (*Introduction*) and chapter five (*Discussion*) for the details.

Support from Learners' Perceptions and Attitudes

Questionnaire and face-to-face semi-structured interviews

The third research question asked about participants' opinions/viewpoints and attitudes towards the simultaneous multimedia glossing modes. The question also sought learners' preference on different glossing modes and the rationale for their choice. The findings of the questionnaire and face-to-face semi-structured interviews showed that the dual glossing modes of L2 definition and audio glossing as well as L2 definition and video/animation glossing were more favorable than the single glossing mode of L2 definition alone in assisting the participants to acquire the words, and keep them in short and long-term memory. However, among the two dual modes, L2 definition and video/animation glossing was the preferred one.

In short, 94% percent of the learners believed that the presentation mode of L2 definition and video/animation facilitated their word learning; about 85% said that the combination of video/animation glossing with L2 definition aided them to recall the new words better in long-term, and more than half of the participants (65%) were eager to use this mode in future for their vocabulary learning. The second favorable glossing mode that made the word learning an easy experience was L2 definition and audio glossing mode, stated by 76% of the participants. This mode also helped 64% of the learners to recall words in long-term, and was preferred by more than half of them (64%). The glossing mode that ranked last was single mode of L2 definition alone. Few participants (less than 10%) believed that learning words via the L2 definitions could be easy (as compared to the other two modes); less than half of the learners (31%) regarded this strategy as a means to help them to remember the words in long-term; and some learners (13%) wanted to use it for future.

Overall, the findings of the questionnaire implied that even though the majority of the participants (75%) favoured the combination of textual definition and visual annotations as their preferred mode, and were considered as the visualizers in this survey, the other 25% of the learners preferred verbal modes of the instruction such as L2 definition alone or L2 definition and audio glossing, and were considered as the verbalizers in this survey. Thus, it is important to attend to the learning preferences and styles (Rassaei, 2017; Plass et al., 1998) of all the learners in a class when instructing the words via multimedia glossing modes.

The face-to-face semi-structured interview also confirmed the findings above upon the favorable mode of L2 definition and video/animation glossing. The rationale of the learners for their choice was that video/animation helped them to keep the words in their mind for later use, and was regarded as an interesting, motivating, and important practice for the comprehension of the texts (Sakar & Erçetin, 2005; Al-Seghayer, 2001), as well as being a very enjoyable word learning experience (Erçetin, 2003).

Pedagogical Implications

The present study has pedagogical implications for the students, teachers, syllabus designers, and materials developers. As shown, the findings revealed that glossing was useful for enhancing L2 learners' word learning and retention. Besides, in general, dual glossing modes encouraged vocabulary learning and recall more effectively than single glossing mode. The following points should be taken into consideration when developing, using, and presenting multimedia teaching materials including glosses in language classrooms:

1. The simultaneous presentation of word definitions accompanied by relevant video/animations or audio recordings facilitates vocabulary learning, and minimizes any extraneous cognitive load on the learners; thus "enhances cognitive processing of multimedia information" (Türk & Erçetin, 2014, p. 16) for long-term word retrieval. Material and syllabus designers as well as teachers who are interested in preparing their own supplementary materials based on their learners' needs should consider the temporal contiguity principle of multimedia learning. The principle posits that presenting verbal and visual multimedia information simultaneously decreases cognitive load; and thus, enhances learning. Additionally, Mayer (2001) recommends that other multimedia principles[69] (e.g., coherence, spatial contiguity, modality, signaling, etc.) be considered when designing hypermedia materials in multimedia-based learning environment.

2. The findings warrant attention in ways to present multimedia information in different glossing combinations and locations. This study only considered pop-up multimedia gloss presentation in that the hyperlinked target word was shown up in three different annotation modes by clicking. However, course book and material designers as well as teachers should take into account both in-text[70], and marginal gloss[71] display modes as well.

---

[69] See Mayer (2014) for a detailed discussion on types of multimedia principles.
[70] Embedding the glosses within the passage next to the target word (Chen, 2016).
[71] Listing the definitions of the target words at the end of the text, page or in the margin (Arpaci, 2016; Chen, 2016).

3. Using multimedia instruments inside language classrooms require the teachers to train and familiarize the students with computer and multimedia software applications in advance. Although the participants of the present study had no choice over the selection of the multimedia materials individually, it is suggested that teachers encourage learners to access the type of multimedia information while learning vocabulary in other multimedia-based learning environments, and assist them to use the multimedia annotations properly (Türk & Erçetin, 2014).

4. This study aimed to encourage learners to expand their vocabulary reservoir and utilize them in long-term by intentionally/explicitly exposing them to target words by means of different glossing modes. However, depending on the learning objectives of a language course in terms of vocabulary learning, appropriate strategies and approaches should be taken into consideration. If the final objective of the course is to acquire the words, and recall them in long-term, then intentional vocabulary learning is welcomed followed by enough word-focused activities and exercises (Laufer & Rozovski-Roitblat, 2015; Schmitt, 2008) so that learners can establish form-meaning relationship, guess the words correctly, and reinforce vocabulary learning and retention (Yoshii, 2014). Thus, to consolidate the link, glosses can act as a mediator to provide learners with repeated exposures and maximum amount of engagement.

5. In preparing and designing multimedia annotations, material developers should be cautioned to select and include annotations that have the same quality and cultural appropriateness. In the present study, attempts were made to choose the video/animation clips based on the available resources such as YouTube, Vimeo, and Daily motion educational websites. However, it is recommended that other resources such as children's cartoons and motion picture dictionaries be utilized.

6. Material developers and teachers should consider the extent to which video/animation clips could be useful, practical, and applicable in the class before applying them. Considering the amount of time spent on designing the syllabus of the course before the class, and the time restrictions of the class itself, preparing motion pictures and animations for all the target words would be time-consuming. Also, it is difficult to select the clips that surely represent the same definition derived from the new words. In other words, although video/animation clips might be one fascinating means of vocabulary teaching/learning to attract learners' attention to a large extent, it is possible that finding the appropriate clips, preparing and adjusting them to the needs of the learners, and fitting them to the requirements of the course might be a time-consuming task, requiring hours of preparation.

7. In designing vocabulary courses/lessons, material developers should keep in mind that selecting multimedia animations and videos for word glossing might be restricted to a series of concrete nouns, action verbs, and some limited adjectives. In addition, instructing other parts of speech such as adverbs, abstract nouns, and stative verbs (i.e., to be) via appropriate video/animation films would be difficult.

8. Vocabulary-based printed materials, instructions, and in-class lessons should be developed to foster students' word retention by means of implementing pictures, graphics, and icons as well as audios/sounds and video/animations.

9. In developing vocabulary materials for instruction (e.g., word-focused activities and exercises), material developers and language teachers should consider individual learners' learning differences (i.e., visualizer-verbalizer dichotomy). Plass and colleagues (1998) state that "visualizer-verbalizer dimension describes individual differences among students when they acquire and process visual versus verbal information" (p. 27).

Research Implications: A Direction to Future Research

This study presented an effort to empirically investigate the effectiveness of multimedia glossing on vocabulary learning in terms of both short and long-term word retention. Three glossing modes, embedded in multimedia texts, were compared. Future studies are required for a comprehensive understanding and confirmation of the appropriateness of multimedia glossing for word learning and recall. The following gives directions to future research on this domain:

1. The present study utilized printed textual definitions, audio recording, and video/animation to showcase the definition of the target words in a multimedia setting. Future research could add static/still-pictures to this combination in order to investigate the impact of verbal and visual glossing types on L2 vocabulary learning.

2. This research examined the effectiveness of multimedia glossing on L2 learners' vocabulary learning and retention at intermediate proficiency level using L2 word definitions. It also considered a pop-up/hyperlink glossing format for the inclusion of glossed words throughout the text. However, replicating the study with participants of different language levels (i.e., beginners, advanced) could yield different results. Research has shown some inconsistencies regarding the impact of glossing modes on L2 learners with different proficiency levels which warrant further research. The low-proficient learners enjoyed glossing in Apraci's (2016) study; whereas high proficient participants performed significantly well in Ko's (2005) research. Also, low proficient learners benefited more from in-text glossing while marginal glossing was more effective for high proficient participants (Yeung, 1999). Additionally, the multimedia text could be designed with different glossing formats and locations (i.e., marginal or in-text glossing).

3. The English language was used as the participants' L2 to provide the definitions of the target words. However, future studies could focus on learners' first language (L1) for vocabulary definitions using similar design. Ko (2012) mentions that limited studies have addressed the inclusion of L1-L2 comparisons for the effect of glossing on vocabulary learning. However, for those with such a comparison, the findings are still mixed and inconsistent. Some studies found no significant differences between L1 and L2 glosses (Chen, 2002; Jacobs et al., 1994; Ko, 2012; Yoshii, 2006); whereas, others revealed the outperformance of L2 gloss for higher proficiency level learners, and L1 gloss for lower-proficiency levels (Miyasako, 2002). Yoshii (2006) also found a significant drop on the delayed post-test of L2 gloss group and no drop for L1 gloss group.

4. Although Mayer (2001) states that "all multimedia messages are not equally effective" (p. 79) for all language learners, it is hoped that replicating the methods and research design of the present study yields interesting findings upon the use of multimedia glossing for vocabulary learning for languages other than English, and with students from countries other than Iran.

5. The focus of the present study was to examine vocabulary learning by instructing the target words intentionally through different glossing modes without considering learners' performance on reading comprehension. Yet, future research can consider incidental/implicit vocabulary learning by exposing learners to authentic reading materials without notifying them of any subsequent vocabulary measurements or compare the two approaches (i.e., incidental vs. intentional), and gauge learners' reading comprehension skill as another variable.

6. Including other variables such as participants' different learning styles and preferences such as visualizers and verbalizers (Plass et al., 1998), and investigating how their learning behaviour differs would lead to different findings. The visualizer-verbalizer dimension is related to "individual differences among students when they acquire and process visual versus verbal information" (Plass et al., 1998, p. 27). In their study, Plass and colleagues found that visualizers benefit from visual representations of multimedia materials whereas verbalizers prefer verbal modes, suggesting the importance of considering the individual differences for vocabulary learning with media types.

7. The present study considered the simultaneous presentation of glossing modes as one means of vocabulary technique for word learning and retention, without considering the successive presentation of glossing modes. Thus, it would be an open area for future research on this domain to investigate and compare the efficacy of display gloss representations on vocabulary learning and retention through both simultaneous and successive conditions.

8. Comparing traditional (paper-based) and multimedia (computer-based) glossing mode conditions could also be a focus of future studies because the two-learning environments may yield interesting and different results. Traditional gloss conditions refer to when the target words are glossed in the margin, along with paper-based images or within the text; whereas, hypertext glossed conditions are when the target words are hyperlinked with images and illustrations, icons, audios, and animations. Investigating learners' vocabulary performance in these two learning conditions may augment insights into the current knowledge of gloss studies.

9. The inclusion of audio glossing in this study provided the pronunciation of the target words by a native speaker. Further research can be conducted to include the articulation of the spellings of the words in order to help them with the correct spelling when required.

10. This study was conducted in an EFL mono-lingual context where the learners shared Farsi as their first language. However, it is suggested that the study be replicated in an ESL multicultural context where participants, coming from different cultural and linguistic backgrounds, incorporate multimedia instruments for their vocabulary learning and retention.

11. This study chose concrete nouns as the target words to be glossed in the texts due to the ease at which video/animations could be found. It is suggested that other parts of speech like verbs and adjectives be incorporated using multimedia-based learning environment rather than nouns alone.

12. This study was conducted with adult participants ranging from 16-25 years old. It would be interesting to consider participants of varying ages such as children to investigate the efficacy of multimedia glossing on their vocabulary learning and retention. Children with lower cognitive abilities and different learning characteristics (Acha, 2009) may have different learning outcomes.

Study Limitations

There were several restrictions posed on this study that follow:

1. The participants of the present study had no control over clicking the glossed words presented to them in each reading passage. In other words, the annotated texts were displayed through one large screen in front of them in a classroom setting with the instructor controlling the time of the gloss presentation in the assigned mode; however, if each individual had an access to the multimedia text on a computer screen in the language laboratory with the possibility of clicking on the words several times and the selection of the mode of gloss annotations (upon learners' request), the results could have been different.

2. The participants were exposed to the non-traditional hypertexts where the target glossed words were hyperlinked with different multimedia glossing modes, but were evaluated with traditional methods of paper and pencil. However, the results might have been different if on-screen tests (i.e., multimedia-based screen) were utilized with the features of both verbal and visual representations for each test item.

3. The study assigned the time constraint of 15-20 minutes for the completion of the PR test, following the results of the pilot study. Nevertheless, if the time allocation was extended, students could complete the test items in their own time, affecting their performance. Research has shown that productive tests require more mental processing for the learners to retrieve the words and produce them; whereas it is probable that learners rely on guessing power to find the correct answer in MC vocabulary measurements. Thus, results could be different if more timing was allotted for the completion of productive vocabulary measurements.

4. This study was conducted in a classroom setting with a large screen and two medium- sized speakers. The hypertexts were presented through the monitor connected to a central computer and a television. However, if the participants had received the instruction in a laboratory where each individual was sitting in a cabin equipped with a headset and a monitor in front of them, their performance might have changed, and the efficacy of multimedia glossing modes might have been different.

5. This study selected concrete nouns as the target words to gloss. Attempts by the researcher were made to find the video/animation clips that clearly demonstrated the exact definitions of the designated words. However, due to the time restriction to keep all the clips at the same length (7-10 seconds)[72], I had to cut

---

[72] See chapter three (*Methodology*) for the relevant discussion.

and edit some sections of the video/animation clips, which might have resulted in less clarity and comprehensibility of the clips.

6. The present study defined two weeks as the long-term due to the time constraint for data collection; however, the effectiveness of multimedia glossing could be examined over longer time span (e.g., one month).

Concluding Remarks: Filling the Gaps in Literature

With the aim of investigating the effectiveness of simultaneous multimedia glossing on L2 learners' vocabulary learning and retention over short and long-term memory, the present study was situated within the two theoretical frameworks of dual-coding theory (Paivio, 1986) and cognitive theory of multimedia learning (Mayer, 2014, 2007). The research sought to respond to the research questions relating to the efficacy of glossing, and different glossing modes. The effectiveness of glossing was investigated through in-sum test comparisons, and the impact of different glossing modes on word learning and retention was measured via sub-test comparisons. Participants' scores on vocabulary tests were considered from pre-to immediate post-tests for short-term effect, and from pre-to delayed post-tests for long-term retention for both of the comparisons. Also, the participants' attitudes and perceptions were evaluated via a questionnaire and face-to-face semi-structured interviews. The findings of the study were supported by the relevant current literature on the domain of multimedia glossing and vocabulary acquisition as well as word retention.

This study is significant as it has provided insight crucial to vocabulary learning and retention. The research filled the gap in the current literature despite the limitations in the following ways: **(a)** the study tried to address the inconclusive and insufficient evidence regarding what gloss combination(s) is more effective in facilitating vocabulary learning and enhancing long-term word recollection. The general findings showed that bimodal glossing formats are more effective than single glossing type to promote learners' vocabulary acquisition, and enhance their long-term word retention. However, the role of single mode of glossing should not be overlooked as it also helped the participants to both learn and retrieve the words in a few test sessions; **(b)** the study showed that, among the dual glossing modes in a computerized learning environment, the combination of text/L2 definition and video/animation glossing was preferred and more efficient than text/L2 definition and audio glossing. The reason lay on the fact that with the audio combination, the learners were exposed to two forms of verbal representations without receiving any visual clues. Thus, the audio mode along with the word definition of the target words may have resulted in the redundancy effect, and affected the performance of the participants; and finally, **(c)** the design of the study as well as its methodological approaches was another addition to the field in that it used mixed methods research to examine the efficacy of multimedia glossing through examining learners' scores on pre/immediate and post-tests as well as their perceptions and attitudes towards glossing modes.

REFERENCES

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Le*arning, *21*(3), 199-226.

AbuSeileek, A. (2008). Hypermedia annotation presentation: learners' preferences and effects on EFL reading comprehension and vocabulary acquisition. *CALICO Journal*, *25*(2), 1-15.

AbuSeileek, A. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computer & Education, 57*(1), 1281-1291.

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology*, *40*(1), 23-31.

Ågerfalk, J. Pär. (2013). Embracing Diversity through Mixed Methods Research. *European Journal of Information Systems*, *22*, 251–256.

Ahangari, S., & Abdollahpour, Z. (2010). The effect of multimedia annotations on Iranian EFL learners' L2 vocabulary acquisition. *The Journal of Applied Linguistics*, *3*(1), 1-18.

Ahmad, J. (2011). Intentional vs. Incidental Vocabulary Learning. *Interdisciplinary Journal of Contemporary Research Business, 3*(5), 67-75.

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

Alessi, S., & Dwyer, A. (2008). Vocabulary assistance before and during reading. *Reading in a Foreign Language*, *20*(2), 246-263.

Al-Ghafli, M. H. (2011). The Effect of Mediated Glosses on Vocabulary Retention and Reading Comprehension with Learners in Saudi Arabia. *Unpublished PhD Dissertation*. University of Kansas.

Al-Seghayer, K. (2001). The Effect of Multimedia Annotation Modes on L2 Vocabulary Acquisition: A comparative study. *Language Learning & Technology, 5*(1), 202-232.

Al-Seghayer, K. (2003). Technological and Pedagogical Consideration for a more Effective Electronic Glossary. *The Reading Matrix, 3*(1), 1-15.

Al-Seghayer, K. (2016). Celebrating the Story of my First Contribution to CALL. *Language Learning & Technology*, *20*(2), 179-184.

Amiryousefi, M., & Ketabi, S. (2011). Mnemonic instruction: A way to boost vocabulary learning and recall. *Journal of Language Teaching and Research*, *2*(1), 178-182.

Angouri, J. (2010). Quantitative, qualitative or both? Combining Methods in Linguistic Research. In L. Litosseliti (ed.), *Research Methods in Linguistics* (pp. 29-45). London: Continuum International Publishing Group.

Appleton, J. V. & King, L. (2002). Journeying from the Philosophical Contemplation of Constructivism to the Methodological Pragmatics of Health Services Research. *Journal of Advanced Nursing*, *40*(6), 641-648.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (1998). *Critical Realism: Essential Readings (Critical Realism: Interventions)*. New York; London: Routledge.

Ariew, R., & Ercetin, G. (2004). Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning*, *17*(2), 237–259.

Arpacı, D. (2016). The effects of accessing L1 versus L2 definitional glosses on L2 learners' reading comprehension and vocabulary learning. *Eurasian Journal of Applied Linguistics, 2*(1), 15–29.

Azari, F. (2012). Review of effects of textual glosses on incidental vocabulary learning. *International Journal of Innovative Ideas, 12*(2), 13-24.

Babaie, S. H., & Razmjoo, S. A. (2015). Multimedia annotation: Comparability of gloss modalities and their implications for reading comprehension. *Journal of English Language Teaching and Learning, 15*, 1-41.

Baddeley, D. (2002). Is Working Memory Still Working? *European Psychologist, 7*(2), 85–97.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of Foreign Language Vocabulary and the Spacing Effect. *Psychological Science, 4*(5), 316–321.

Barcroft, J. (2009). Effects of Synonym Generation on Incidental and Intentional L2 Vocabulary Learning during Reading. *TESOL Quarterly*, *43*(1), 79-103.

Bell, F. L., & LeBlanc, L. B. (2000). The Language of Glosses in L2 Reading on Computer: Learners' Preferences. *Hispania, 83*(2), 274-285.

Black, A., Wright, P., Black, D., & Norman, K, (1992). Consulting on-line dictionary information while reading. *Hypermedia*, *4*(3), 145-169.

Boers, F., Warren, P., He, L., & Deconinck, J. (2017). Does Adding Pictures to Glosses Enhance Vocabulary Uptake from Reading? *System*, *66*, 113-129. Retrieved online on 10 May 2017 from http://www.sciencedirect.com/science/article/pii/S0346251X17302415.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening, 14*, 14–31.

Boonkongsaen, N. (2012). Factors Affecting Vocabulary Learning Strategies: A Synthesized Study. *Naresuan University Journal, 20*(2), 45-53.

Bowles, M. A. (2004). L2 glossing: To CALL or not CALL. *Hispania, 87*(3), 541-552.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental Vocabulary Acquisition from Reading, Reading-while-listening, and Listening to stories. *Reading in a Foreign Language, 20*(1), 136-163.

Brundrett, M, & Rhodes, C. (2014). *Researching educational leadership and management: Methods and approaches*. Los Angeles; London; New Delhi; Singapore; Washington, DC: Sage.

Bryman, A. (1992). Quantitative and Qualitative research: further reflections on their integration. In J. Brannen (ed.), *Mixing Methods: Qualitative and Quantitative Research* (pp. 57-81). Aldershot Hants: Avebury.

Bryman, A. (1998). Quantitative and Qualitative Research Strategies in Knowing the Social World. In T. May & M. Williams (eds.), *Knowing the Social World* (pp. 138–157). Buckingham: Open University Press.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*(1), 182-197.

Butcher, K. R. (2014). The Multimedia Principle. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning* (pp. 174-206). New York: Cambridge University Press.

Cabeza, R., Kapur, S., Craik, F. I. M., McIntosh, A. R., Houle, S., & Tulving, E. (1997). Functional neuroanatomy of recall and recognition: A PET study of episodic memory. *Journal of Cognitive Neuroscience, 9*(2), 254-265.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior, 26*(4), 413–437.

Celce-Murcia, M. (2001). Language teaching approaches: An overview. In M. Celce-Murcia, (ed.), *Teaching English as a second or foreign language* (pp. 3-11). Boston, MA: Heinle & Heinle.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition & Instruction*, *8*(4), 293-332.

Chen, Z. (2006). The effects of multimedia annotations on L2 vocabulary immediate recall and reading comprehension: A comparative study of text-picture and audio-picture annotations under incidental and intentional learning conditions. *Unpublished PhD Dissertation.* Department of World Language Education College of Arts and Sciences, University of South Florida. Retrieved online on 1 September 2016 from http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=3477&context=etd

Chen, C., & Truscott, J. (2010). The Effects of Repetition and L1 Lexicalization on Incidental Vocabulary Acquisition. *Applied Linguistics, 31*(5), 693–713.

Chen, I. J., & Yen, J. C. (2013). Hypertext Annotation: Effects of Presentation Formats and Learner Proficiency on Reading Comprehension and Vocabulary Learning in Foreign Languages. *Computers & Education*, *63*, 416-423.

Chen, I. (2016). Hypertext glosses for foreign language reading comprehension and vocabulary acquisition: effect of assessment methods. *Computer Assisted Language Learning*, *29*(2), 413-426.

Cheng, Y., & Good, R. L. (2009). Glosses: Effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language, 21*(2), 119-142.

Choi, S., Kim, J., & Ryu, K. (2014). Effects of context on implicit and explicit lexical knowledge. *Neuropsychologia*, *63*, 226–234.

Choi, S. (2016). Effect of L1 and L2 Glosses on Incidental Vocabulary Acquisition and Lexical Representations. *Learning and Individual Differences*, *45*, 137-143.

Chun, D. M., & Plass, J. L. (1996a). Effects of Multimedia Annotations on Vocabulary Acquisition. *The Modern Language Journal, 80*(2), 183-198.

Chun, D. M., & Plass, J. L. (1996b). Facilitating Reading Comprehension with Multimedia. *System*, *24*(4), 503-519.

Chun, E., Choi, S., & Kim, J. (2012). The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study. *Neuroscience Letters*, *521*(2), 125-129.

Clark, J. M., & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review, 3*(3), 149-210.

Coleman, M., & Liau, T. L. (1975). A Computer Readability Formula designed for Machine Scoring. *Journal of Applied Psychology*, *60*, 283-284.

Colosi, L. (2006). Designing an effective questionnaire. Cornell Cooperative Extension. Retrieved online on 1 September 2016 from http://www.human.cornell.edu/pam/outreach/parenting/research/upload/Designing-20an-20Effective-20Questionnaire.pdf

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Approaches*. Thousand Oaks, CA: Sage.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and Conducting Mixed Methods Research*, (2nd edition). Thousand Oaks, CA: Sage Publications.

Danermark, B., Ekstrom, M., Jakobsen, L., & Karlsson, J. C. (2002). *Explaining Society: Critical Realism in the Social Sciences*. London: Routledge.

Denzin, N. K. (1970). *The Research Act: A Theoretical Introduction to Sociological Methods*. Chicago: Aldine Publishing Company.

Denzin, N. K. (1978). *The Research Act: A theoretical Introduction to Sociological Methods*. New York: McGraw-Hill.

De Ridder, I. (2002). *Reading from the screen in a second language*. Apeldoorn: Granat Publishers.

Devi, A. V. (2010). *Computer Assisted Language Learning Teaching and Testing (CALLTT): Computer assisted language teaching and testing with multimedia material – some aspects*. Saarbrucken: LAP LAMBERT Academic.

Eckerth, J., & Tavakoli, P. (2012). The Effects of Word Exposure Frequency and Elaboration of Word Processing on Incidental L2 Vocabulary Acquisition through Reading. *Language Teaching Research, 16*(2), 227-252.

Erçetin, G. (2003). Exploring ESL learners' Use of Hypermedia Reading Glosses. *CALICO Journal, 20*(2), 261-283.

Elley, W. B. (1989). Vocabulary Acquisition from Listening to Stories. *Reading Research Quarterly, 24*(2), 174-87.

Ellis, N.C. (1997). Vocabulary acquisition: word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (eds.), *Vocabulary acquisition, word structure, collocation, word class, and meaning* (pp. 122-139). Cambridge, UK: Cambridge University Press.

Elis, R. (2006). Current Issues in the Teaching of Grammar: An SLA Perspective. *TESOL Quarterly, 40*(1), 83-107. doi: 10.2307/40264512.

Farvardin, M., & Biria, R. (2012). The Impact of Gloss types on Iranian EFL Students' Reading Comprehension and Lexical Retention. *International Journal of Instruction, 5*(1), 99-114.

Field, A. P. (2009). *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll* (3rd edition). London: Sage publications.

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics* (4th edition). London: Sage Publications.

File, K. A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly, 44*(2), 222-249.

Fletcher, J. D., & Tobias, S. (2005). The Multimedia Principle. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning* (pp. 117-133). New York, NY: Cambridge University Press.

Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology, 32,* 221-233.

Foddy, W. (1993). *Constructing Questions for Interviews and Questionnaires*. New York, NY: Cambridge University Press.

Fontana, A., & Frey, J. H. (1994). Interviewing: The Art of Science. In N. K. Denzin & Y. S. Lincoln (eds.), *The Handbook of Qualitative Research* (pp. 361-76). Thousand Oaks: Sage Publications.

Fontana, A., & Frey, J. H. (2005). The interview: From neutral stance to political involvement. In N. K. Denzin & Y. S. Lincoln (eds.), *The Sage handbook of qualitative research* (3rd edition) (pp. 695-727). Thousand Oaks, CA: Sage Publications, Inc.

Gass, S. (1999). Discussion: incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*(2), 319-333.

Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational Research: Competencies for Analysis and Applications* (9th ed.). Toronto: Pearson Education.

Gettys, S., Imhof, L. A., & Kautz, J. O. (2001). Computer Assisted Reading: the effect of glossing format on comprehension and vocabulary retention. *Foreign Language Annals*, *34*(2), 91-99.

Ghahari, S, & Heidarolad, M. (2015). Multiple-choice Glosses and Incidental Vocabulary Learning: A Case of EFL Context. *Reading Matrix*, *15*(1), 262-273.

Gardner, D. (2011). *Fostering autonomy in language learning*. Gaziantep: Zirve University.

Goodbridge, V. (2010). Why Vocabulary is so important? *Memory, Mnemonics, and the lexicon*. IH Barcelona Annual Conference.

Ghorbani, M., & Karimi, N. (2011). The Impact of Memory Strategy Instruction on Learners' EFL Vocabulary Retention. *Theory and Practice in Language Studies, 1*(9), 1222-1226.

Grabe, W., & Stoller, F. L. (1997). Reading and Vocabulary Development in a Second Language: A Case Study. In J. Coady, J. Coady & T. Huckin (eds.), *Second*

Language Vocabulary Acquisition: A rationale for pedagogy (pp. 99-122). Cambridge, UK: Cambridge University Press.

Greene, J. C. (2007). *Mixed Methods in Social Inquiry*. San Francisco: Jossey-Bass.

Groot, P. (2000). Computer Assisted Second Language Vocabulary Acquisition. *Language Learning & Technology*, *4*(1), 60-81.

Gu, Y. (2003). Vocabulary Learning in a Second Language: Person, Task, Context and Strategies. *TESL-EJ*, *7*(2). Retrieved online on 18 September 2014 from http://www-writing.berkeley.edu/TESL-EJ/ej26/a4.html.

Gunning, R. (1952). *The Technique for Clear writing*. New York: McGraw-Hill.

Gyllstad, H., Vilkaite, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *International Journal of Applied Linguistics*, *166*(2), 276-303. doi: 10.1075/itl.166.2.04gyl.

Harley, B. (Ed.). (1995). *Lexical Issues in Language Learning*. Ann Arbor, MI: Research Club in Language Learning.

Hartman, P., & Kirn, E. (2014). The Silk Road. In P. Hartman & E. Kirn (eds.), *Interactions Reading 2* (pp. 171-172). Mishawaka, IN: The McGraw-Hills companies.

Hong, X. (2010). Review of Effects of Glosses on Incidental Vocabulary Learning and Reading Comprehension. *Chinese Journal of Applied Linguistics*, *33*(1), 56-72.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a Foreign Language*, *11*(2), 207-223.

Hu, M., & Nation, I. S. P. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, *13*(1), 403-430.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *21*(2), 181-193.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (eds.), *Vocabulary and Applied Linguistics* (pp. 113-125). London: MacMillan.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary use, and Reoccurrence of Unknown Words. *The Modern Language Journal*, *80*(3), 327-339.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *15*(3), 539-558.

Hüllen, W. (1989). In the beginning was the gloss. In G. James (ed.), *Lexicographers and their works* (pp. 100-116). Exeter, UK: University of Exeter.

Hunt, A., & Beglar, D. (1998). Current Research and Practice in Teaching Vocabulary. *The Language Teacher Online, 01*, 258-266. Retrieved online on 01 October 2014 from http://www.jalt-publications.org/old_tlt/articles/1998/01/hunt.

Hunt, A., & Beglar, D. (2005). A Framework for developing EFL Reading Vocabulary. *Reading in a Foreign Language, 17*(1), 23-59.

Hussein, A. (2009). The Use of Triangulation in Social Sciences Research: Can Qualitative and Quantitative Methods be Combined? *Journal of Comparative Social Work, 4*(1), 1-12.

Jacobs, G. (1994). What Lurks in the Margin: Use of Vocabulary Glosses as a Strategy in Second Language Reading. *Issues in Applied Linguistics*, *5*(1), 115-137.

Jacobs, G. M., DuFon, P., & Hong, F. C. (1994). L1 and L2 Vocabulary Glosses in L2 Reading Passages: Their Effectiveness for Increasing Comprehension and Vocabulary Knowledge. *Journal of Research in Reading, 17*(1), 19-28.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm whose Time Has Come. *Educational Researcher, 33*(7), 14-26.

Jones, L. C. (2004). Testing L2 Vocabulary Recognition and Recall using Pictorial and Written Test Items. *Language Learning & Technology*, *8*(3), 122-143.

Jones, L. C. & Plass, J. L. (2002). Supporting Listening Comprehension and Vocabulary Acquisition in French with Multimedia Annotations. *The Modern Language Journal*, *86*(4) 546–561. doi: 10.1111/1540-4781.00160

Jung, J. (2015). Effects of Glosses on Learning of L2 Grammar and Vocabulary. *Language Teaching Research, 19*(2), 1-21. Retrieved online on 1 April 2015 from http://ltr.sagepub.com/content/early/2015/03/12/1362168815571151.abstract.

Jung, J. (2016). Effects of glosses on learning of L2 grammar and vocabulary. *Language Teaching Research, 20*(1), 92-112.

Kim, D., & Gilman, D. A. (2008). Effects of Text, Audio, and Graphic Aids in Multimedia Instruction for Vocabulary Learning. *Educational Technology and Society*, *11*(3), 114-126.

Kim, D., & Kim, D. (2012). Effect of screen size on multimedia vocabulary learning. *British Journal of Educational Psychology*, *43*(1), 62-70.

Khezrlou, S., & Ellis, R. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System*, *65*, 104-116.

Knight, S. (1994). Dictionary Use while Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities. *The Modern Language Journal*, *78*(3), 285-299.

Ko, H. M. (2005). Glosses, Comprehension, and Strategy Use. *Reading in a Foreign Language*, *17*(2), 125-143.

Ko, H. M. (2012). Glossing and Second Language Vocabulary Learning. *TESOL Quarterly*, *46*(1), 56-70.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and Pictorial Glosses: Effectiveness on Incidental Vocabulary Growth when Reading in a Foreign Language. *Foreign Language Annals*, *32*(1), 89-113.

Larson-Hall, J. (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York, NY: Routledge, Taylor & Francis Groups.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (eds.), Special language: From humans thinking to thinking machines (pp. 316–323). Clevedon: Multilingual Matters.

Laufer, B. (1990). Why are Some Words More Difficult than Others? *IRAL: International Review of Applied Linguistics in Language Teaching*, *28*(4), 293-307.

Laufer, B. (1992). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading*, *15*, 95-103.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20-34). Cambridge: Cambridge University Press.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology, 3*(2), 58-76.

Laufer, B., & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-Induced Involvement. *Applied Linguistics*, 22(1), 1-26.

Laufer, B. and Goldstein, Z. (2004), Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning*, *54*(3), 399-436.

Laufer, B. (2005). Focus on Form in second language vocabulary acquisition. In S. H. Foster-Cohen, M. P. Garcia-Mayo & J. Cenoz. Benjamins (eds.) *EUROSLA*

Yearbook 5 (pp. 223-250). Edited by Susan H. Foster-Cohen, María del Pilar García Mayo and Jasone Cenoz. doi: 10.1075/eurosla.5.11lau.

Laufer-Dvorkin, B. (2006). Comparing focus on form and focus on FormS in second-language vocabulary learning. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes, 63*(1), 149-166. doi: 10.1353/cml.2006.0047.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research, 15*(4), 391-411.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research, 16*(9), 687-711.

Lawson, M. J., & Hogben, D. (1996). The Vocabulary Learning Strategies of Foreign Language Students. *Language Learning*, *46*(1), 101-135.

Leffa, V. (1992). Making Foreign Language Texts Comprehensible for Beginners: An Experiment with an Electronic Glossary. *System*, *21*(1), 63-73.

Lenders, O. (2008). Electronic Glossing – is it worth the effort? *Computer Assisted Language Learning, 21*(5), 457-481.

Levene, H. (1960). Robust testes for equality of variances. In I. Olkin (ed.), *Contributions to Probability and Statistics* (pp. 278-292). Palo Alto, CA: Stanford Univ. Press.

Lin, C. (2009). Learning Action Verbs with Animation. *The JALT CALL Journal*, *5*(3), 23-40.

Lin, C., & Tseng, Y. (2012). Videos and Animations for Vocabulary Learning: A study On Difficult words. *TOJET: The Turkish Online Journal of Educational Technology*, *11*(4), 346-355.

Liu, T. & Lin, P. (2011). What comes with technological convenience? Exploring the behaviors and performances of learning with computer-mediated dictionaries. *Computers in Human Behavior*, *27*(1) 373-383.

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, *19*(1), 1-14.

Lomicka, L. L. (1998). To gloss or not to gloss: An investigation of Reading Comprehension on line. *Language Learning and Technology, 1*(2), 41-50.

Lyman-Hager, M. A., & Davis, J. N. (1996). The Case for Computer-Mediated Reading: Une Vie de Boy. *The French Review, 69*(5), 775-792.

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, *43*(2), 263-287.

Marefat, H., Rezaee, A.A., & Naserieh, F. (2016). Effect of computerized gloss presentation format on reading comprehension: A cognitive load perspective. *Journal of Information Technology Education: Research*, *15*, 479-501. Retrieved online on 23 November 2016 from http://www.informingscience.org/Publications/3568.

Martin, W. E., & Bridgmon, K. D. (2012). *Quantitative and Statistical Research Methods - from hypothesis to results*. Somerset, NJ: Wiley.

Marzban, A., (2011). Investigating the Role of Multimedia Annotations in EFL Reading Comprehension. *Procedia-Social and Behavioral Sciences*, *28*, 72-77.

Mayer, R. E. (1997). Multimedia Learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

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

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31-49). New York: Cambridge University Press.

Mayer, R. E. (2005a). Cognitive Theory of Multimedia Learning. In R. E. Mayer (ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31-48). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004.

Mayer, R. E. (2005b). Cognitive Theory of Multimedia Learning. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.

Mayer, R. E. (2008). Applying the Science of Learning: Evidence-Based Principles for the Design of Multimedia Instruction. Journal of American Psychologist, *63*(8), 760-769.

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

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, *83*(4), 484-490.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing Constructivist Learning from Multimedia Communications by Minimizing Cognitive Load. *Journal of Educational Psychology*, *91*(4), 638-643.

Mayer, R.E., & Fiorella, L. (2014). Principles of Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning* (pp. 279-315). New York: Cambridge University Press.

Mayer, R. E., & Moreno, R. (2002). Aids to Computer-Based Multimedia Learning. *Journal of Learning and Instruction*, *12*, 107-199.

Mayer, R., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1). 42-52.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389-401.

Mayer, A. (2013). *Introduction to Statistics and SPSS in Psychology*. Harlow: Pearson Education.

Mayer, R. E. (ed.). (2014). *The Cambridge handbook of Multimedia Learning*. New York: Cambridge University Press.

McAllister, D., & Guidice, R. M. (2012). This is only a test: a machine-graded improvement to the multiple-choice and true-false examination. *Teaching in Higher Education*, *17*(2), 193-207.

McCarthy, M. (1984). A new look at Vocabulary in EFL. *Applied Linguistics, 5*(1), 12-27.

McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.

McLaughlin, G. (1969). SMOG Grading-A New Readability Formula. *Journal of Reading*, *12*(8), 639-646.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*(2), 142-154.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An extended sourcebook* (2nd edition). Thousand Oaks: Sage Publications.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rdedition). Los Angeles; London; New Delhi; Singapore; Washington, DC: Sage.

Milton, J. (2008). French Vocabulary Breadth among Learners in the British school and University System: Comparing Knowledge over Time. *Journal of French Language Studies*, *18*(3), 333-348.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.

Min, H. T. (2008). EFL Vocabulary Acquisition and Retention: Reading plus Vocabulary Enhancement Activities and Narrow Reading. *Language Learning, 58*(1), 73-115.

Minichiello, V., Aroni, R., Timewell, E., & Alexander, L. (1990). *In-depth Interviewing: Researching People*. Hong Kong: Longman Cheshire Pty Limited.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology*, *39*, 1-20.

Mizumoto, A., & Kansai, O. T. (2009). Examining the Effectiveness of Explicit Instruction of Vocabulary Learning Strategies with Japanese EFL University Students. *Language Teaching Research, 13*(4), 4275-449.

Mohsen, M. A. (2011). Effects of Multimedia Glosses on Aiding Vocabulary Acquisition in EFL Environment. *Language in India*, *11*(3), 278-302.

Mohsen, M. A., & Balakumar, M. (2011). A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL*, *23*(02), 135-159.

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

Moreno, R., & Mayer, R. E. (2000a). A Coherence Effect in Multimedia Learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, *92*(1), 117-125.

Moreno, R., & Mayer, R. E. (2000b). Engaging Students in Active Learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 92*(4), 724-733.

Nagata, N. (1999). The Effectiveness of Computer Assisted Interactive Glosses. *Foreign Language Annals*, *32*(4), 469-479.

Nagy, W. E. (2005). Why Vocabulary Instruction Needs to Be Long-Term and Comprehensive. In E. H. Hiebert & M. L. Kamil (eds.), *Teaching and Learning Vocabulary Bringing Research to Practice* (pp. 27-44). Mahwah, NJ: Erlbaum.

Najjar, L. J. (1996). Multimedia Information and Learning. *Journal of Educational Multimedia and Hypermedia, 5*(2), 129-150.

Nation, I. S. P. (1983). *Teaching and Learning Vocabulary*. Wellington: English Language Institute, Victoria University.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.

Nation, I. S. P. (1993). Using Dictionaries to Estimate Vocabulary Size: Essential, but rarely followed, procedures. *Language Testing, 10*(1), 27-40.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. New York: Cambridge University Press.

Nation, I. S. P. (2006). How Large a Vocabulary is needed for Reading and Listening? *Modern Language Review, 63*(1), 59-82.

Nation, I. S. P. (2007) Vocabulary Learning through Experience Tasks. *Language Forum, 33*(2), 33-43.

Nation, I. S. P. (2010). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2011). Research into Practice: Vocabulary. *Language Teaching, 44*(4), 529-539.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd edition). Cambridge: Cambridge University Press.

Nation, I.S.P. & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston, MA: Heinle.

Nemati, A. (2009). Enhancing Long-term Retention by Memory Vocabulary Learning Strategies. *Asia TEFL Journal, 6*(4), 14-24.

Nemati, A. (2013). Vocabulary Learning Strategies: A Short Way to Long Term Retention. *Linguistics and Literature Studies*, *1*(1), 8-14.

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, *17*(2), 164–187.

NICHD, National Reading Panel. (2000). *Teaching Children to Read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.

Nunnally, J. O. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Olsen, W. (2004). Triangulation in Social Research: Qualitative and Quantitative Methods can Really Be Mixed. In M. Holborn (ed.), *Developments in Sociology* (pp. 103-118). Ormskirk: Causeway Press.

Oxford, R. L. (1990). *Language Learning Strategies: what every teacher should know*. Boston: Heinle and Heinle Publishers.

Oxford, R., & Crookkall, D. (1990). Vocabualry learning: A critical analysis of techniques. *TESL Canada Journal/Revue TESL du Canada*, *7*(2), 9-30.

Öztürk, M. (2007). Multiple-Choice Test Items of Foreign Language Vocabulary. *Eğitim Fakültesi Dergisi, XX* (2), 399-426.

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

Paivio, A. (1986). *Mental Representations: A Dual-Coding Approach*. New York: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255-287.

Paivio, A. (2007). *Mind and Its Evolution: A Dual Coding Theoretical Approach*. Mahwah, NJ: L. Erlbaum Associates.

Pak, J. (1986). The effect of vocabulary glossing on ESL reading comprehension. Unpublished manuscript, University of Hawaii at Manoa.

Pallant, J. (2013). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS* (4th ed.). Crows Nest, NSW: Allen & Unwin.

Paribakht, T.S. & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition: A rationale for pedagogy,* (pp. 174-199). Cambridge: Cambridge University Press.

Paribakht, T. S., & Wesche, M. (1999). Reading and Incidental L2 Vocabulary Acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition, 21*(2), 195-224.

Pass, F., & Sweller, J. (2014). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (ed.), *The Cambridge handbook of multimedia learning* (pp. 27-42). New York: Cambridge University Press.

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods* (3rd edition). Thousand Oaks, CA: Sage.

Perez, M. M., Peters, E., Calrebout, G., & Desmet, P. (2014). Effects of Captioning on Video Comprehension and Incidental Vocabulary learning. *Language Learning and Technology, 18*(1), 118-141.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners’ form recall of single words and collocations. *Language Teaching Research, 18*(1), 75-94.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting Visual and Verbal Learning Preferences in a Second Language Multimedia Learning Environment. *Journal of Educational Psychology, 90*(1), 25-36.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*, 221-243.

Plass, J., & Jones, L. (2005). Multimedia Learning in Second Language Acquisition. In R. E. Mayer (ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 467-488). New York: Cambridge University Press.

Preston, A. (2007). Memory. *Journal of Scientific America Online*. Retrieved online on 18 June, 2014 from http://www.scienceamerican.com.

Punch, K. F., (2011). *Introduction to Research Methods in Education*. Sage Publication.

Ramezanali, N. (2003). The Effect of Four Different Modes of Instruction on L2 Vocabulary Acquisition of EFL Learners. *Unpublished Master's Thesis*. University of Al-Zahra, Iran.

Rassaei, E. (2017). Computer-mediated Textual and Audio Glosses, Perceptual Style and L2 Vocabulary Learning. *Language Teaching Research*, 1-19. Retrieved online on 20 April, 2017 from http://journals.sagepub.com/doi/abs/10.1177/1362168817690183.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Pres.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, *24*, 146-161.

Reed, S. K. (2010). *Cognition: Theories and application* (8th edition). Belmont, CA: Wadsworth Cengage Learning.

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. Close-ended Questions in Web Questionnaires. *Developments in Applied Statistics, 19*, 159-177.

Riazi, A., & Alvari, A. (2004). Strategy Activation in Learning English Words, *Academic Exchange Quarterly*, *8*(2), 199-203.

Richards, J. C. (2004). *Interchange Student's book* (3rd edition). Cambridge University Press. Cambridge. ISBN 10: 0521601711.

Ridenour, C. S., & Newman, I. (2008). *Mixed Methods Research: Exploring the Interactive Continuum*. Carbondale, IL: Southern Illinois University Press.

Rieber, L. P. (1994). *Computers, Graphics, and Learning*. Madison, WI: WCB Brown & Benchmark.

Roby, W. B. (1999). What's in a Gloss? *Language Learning and Technology, 2*(2), 94-101.

Robey, D. (1996). Research Commentary: Diversity in Information Systems Research: Threat, Promise, and Responsibility. *Information Systems Research*, *7*(4), 400-408.

Rodriguez, M., & Sadowki, M. (2000). Effects of Rote, Context, Keyword and Context/ Keyword Methods on Retention of Vocabulary in EFL Classrooms. *Language Learning*, *50*(2), 385-412.

Rott, S. (1999). The Effect of Exposure Frequency on Intermediate Language Learners' Incidental Vocabulary Acquisition through Reading. *Studies in Second Language Acquisition*, *21*(4), 589-619.

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language, 17*(2), 95-124. Retrieved online on 14 February 2009, from http://nflrc.hawaii.edu/rfl/October2005/rott/rott.pdf

Rott, S. (2007). The effect of frequency of input enhancements on word learning and text comprehension. *Language Learning*, *57*(2), 165-199.

Rott, S., Williams, J., & Cameron, R. (2002). The Effect of Multiple-choice L1 Glosses and Input-output Cycles on Lexical Acquisition and Retention. *Language Teaching Research*, *6*(3), 183-222.

Rouhi, A., & Mohebbi, H. (2013). Glosses, Spatial Intelligence and L2 Vocabulary Learning in Multimedia Context. *The Southeast Asian Journal of English Language Studies*, *19*(2), 75-87.

Rusanganwa, J. A. (2015). Developing a Multimedia Instrument for Technical Vocabulary Learning: a case of EFL undergraduate physics education. *Computer Assisted Language Learning*, *28*(2), 97-111.

Sadeghi, K., & Ahmadi, N. (2012). The Effect of Gloss Type and Mode on Iranian EFL Learners' Reading Comprehension. *English Language Teaching*, *5*(12), 100-110.

Sadeghi, K., Khezrlou, S., & Modirkhameneh, S. (2016). Calling Iranian Learners of L2 English: Effect of Gloss Type on Lexical retention and Reading performance under Different Learning Conditions. *Journal of Research in Reading*, *39*(1), 1-21. doi: 10.1111/1467-9817.12088.

Salem, B. E. (2006). The Influence of Electronic Glosses on Word Retention and Reading Comprehension with Spanish Language Learners. *Doctoral Dissertation*. University of Kansas. Retrieved online on 01 January 2015 from http://www.edtech.ku.edu/research/gloss/publications/bensalem.pdf

Salem, B. E., & Aust, R. (2007). The Influence of Feature-rich Computerized Glosses on Reading Comprehension and Vocabulary Acquisition. *Proceedings of the Sixth Lasted International Conference web-based Education, Volume 2, 663-668*. Chamonix, France. Retrieved online on 22 April 2015 from http://www.edtech.ku.edu/research/papers/IASTED07_Elyes_Aust.pdf**.**

Sakar, A., & Erçetin, G. (2005). Effectiveness of hypermedia annotation for foreign language reading. Journal of Computer Assisted Learning, *21*, 28-38.

Samian, H. V., Foo, T. C. V., & Mohebbi, H. (2016). The Effects of Giving and Receiving Marginal L1 Glosses on L2 Vocabulary Learning by Upper Secondary Learners. *English Language Teaching*, *9*(2), 66-76.

Sato, T. (2016). Applicability of technology-enhanced visual glosses for explicit L2 vocabulary learning: The enhancement of metaphoric competence through the learning of English polysemous words. *Ampersand*, *3*, 43-51. Retrieved online on 17 August 2016 from www.elsevier.com/locate/amper.

Sayer, A. (2000). *Realism and Social Science*. Sage Publications Ltd. London.

Scafidi-Iannone, R. (2012). A Classroom Research Study on Strategies to Promote Vocabulary Retention in the LOTE Classroom. *Doctoral Dissertation*. University of Stony Brook. Retrieved online on 02 November 2014 from https://dspace.sunyconnect.suny.edu/handle/1951/59854.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11*(2), 129-158.

Schmitt, N., & Schmitt, D. (1995). Vocabulary Notebooks: Theoretical Underpinnings and Practical Suggestions. *English Language Teaching Journal, 49*(2), 133-143.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55-88.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.

Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, associations, collocation and word-class knowledge. *Language Testing*, *16*(2), 189-216.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (ed.). (2002). *An Introduction to Applied Linguistics*. Hodder Arnold Press.

Schmitt, N. (2008). Instructed Second Language Vocabulary Learning. *Language Teaching Research, 12*(3), 329-363.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. New York: Palgrave Macmillan Press.

Senter, R.J., & Smith, E.A. (1967). *Automated Readability Index*. Wright-Patterson Air Force Base. P. iii. AMRL-TR-6620. Springfield, Vikkinia. Retrieved online on 01 March 2016 from http://www.dtic.mil/dtic/tr/fulltext/u2/667273.pdf.

Scott, D. (2010). *Education, Epistemology and Critical Realism*. Routledge. New York, NY.

Shahrokni, A. (2009). Second Language Incidental Vocabulary Learning: The Effect of Online Textual, Pictorial, and Textual Pictorial Glosses. *The Electronic Journal for English as a Second Language (TESL-EJ), 13*(3), 1-17.

So, W, Y. (2010). Short-term and long-term retention of new words: Investigating the role of L1 glossing in vocabulary learning among Hong Kong ESL learners. *Unpublished Master's thesis*. University of Hong Kong. Retrieved online on 10 October 2016 from http://dx.doi.org/10.5353/th_b4517659.

Stevick, E. W. (1996). *Memory, Meaning and Method*. (2nd edition). Boston: Heinle and Heinle.

Stewart, J. (2012). A Multiple-choice test of active vocabulary knowledge. *Vocabulary Learning and Instruction*, *1*(1), 53-59.

Stewart, R. A., & Cross, T. L. (1991). The Effect of Marginal Glosses on Reading Comprehension and Retention. *Journal of Reading, 35*(1), 4-12.

Stewart, R. A., & Cross, T. L. (1993). A Field Test of Five Forms of Marginal Gloss Study Guide: An ecological study. *Reading Psychology, 14*(2), 113-139.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312. doi:10.1016/0959-4752(94)90003-5.

Sweller, J., & Chandler, P. (1994). Why Some Material is Difficult to Learn. *Cognition and Instruction*, *12*(3), 185-233.

Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 19-31). New York: Cambridge University Press.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, *22*(2), 123-138.

Sydorenk, T. (2010). Modality of Input and Vocabulary Acquisition. *Language Learning and Technology*, *14*(2), 50-73.

Tabachnick, G. G., & Fidell, L. S. (2007). *Experimental Designs Using ANOVA.* Belmont, CA: Duxbury.

Tabachnick, G. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th edition). Boston: Pearson.

Tabatabaei, O., & Shams, N. (2011). The Effect of Multimedia Glosses on Online Computerized L2 Text Comprehension and Vocabulary Learning of Iranian EFL learners. *Journal of Language Teaching and Research, 2*(3), 714-725.

Tashakkori, A. & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, Calif. Sage Publications. California.

Tashakkori, A., & Teddlie, C. (2003a*). Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks: Sage.

Tashakkori, A., & Teddlie, C. (2003b). The Past and the Future of Mixed Model Research: From "Methodological Triangulation" to "Mixed Model Designs". In A. Tashakkori & C. Teddlie (eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 671-703). Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2008). Quality of Inferences in Mixed Methods Research. In M. Bergman (ed.), *Advances in Mixed Methods Research: Theories and Applications* (pp. 101-119). London, UK: Sage.

Tashakkori, A., & Creswell, J. (2007). Exploring the Nature of Research Questions in Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(3), 207-211.

Tavakoli, M., & Gerami, E. (2013). The Effect of Keyword and Pictorial Methods on EFL Learners' Vocabulary Learning and Retention. *Porta Linguarum, 19*, 299-316.

Teddlie, Ch., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage.

Thaine, C. (2012). Problems in the Natural World. In *Cambridge Academic English B1 Intermediate: an integrated skills course for EAP* (pp. 30-33). Cambridge: Cambridge University Press.

Türk, E., & Erçetin, G. (2014). Effects of Interactive Versus Simultaneous Display of Multimedia Glosses on L2 Reading Comprehension and Incidental Vocabulary Learning. *Computer Assisted Language Learning, 27*(1), 1-26.

Underwood, M. (1989). *Teaching listening*. New York: Longman.

Vahedi, V. S., Ghonsooly, B., & Pishghadam, R. (2016). Vocabulary Glossing: A Meta-analysis of the Relative Effectiveness of Different Gloss Types on L2 Vocabulary Acquisition, *Teaching English with Technology*, *16*(1), 3-25.

Varol, B., & Erçetin, G. (2016). Effects of Working Memory and Gloss Type on L2 Text Comprehension and Incidental Vocabulary Learning in Computer-based Reading. *Procedia-Social and Behavioral Sciences*, *232*, 759-768. International Conference on Teaching and Learning English as an Additional Language, GlobELT 2016, 14-17 April 2016, Antalya, Turkey. Retrieved online on 23 November 2016 from http://dx.doi.org/10.1016/j.sbspro.2016.10.103.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information System. *IS Quarterly*, *37*(1), 21-54.

Wang, Y. (2016). Promoting contextual vocabulary learning through an adaptive computer-assisted EFL reading system. *Journal of Computer Assisted Learning, 32*(4), 291-303. doi: 10.1111/jcal.12132.

Watanabe, Y. (1997). Input, Intake, and Retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition Journal (SSLA), 19*(3), 287-307.

Waring, R., & Takaki, M. (2003). At What Rate do Learners Learn and Retain New Vocabulary from Reading a Graded Reader? *Reading in a Foreign Language, 15*(2), 130-163.

Waring, R. (1997a). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata* [occasional papers of Notre Dame Seishin University, Okayama], *1*, 53-68.

Waring, R. (1997b). A study of receptive and productive learning from word cars. *Studies in Foreign Languages and Literature*, *21*, 94-114.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(1), 33-52.

Webb, S. (2007a). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63-81.

Webb, S. (2007b). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics, 28*(1), 46-65.

Webb, S. (2010). Using Glossaries to Increase the lexical Coverage of Television programs. *Reading in a Foreign Language*, *22*(1), 201-221.

Webb, S., & Chang, A, C.-S. (2015). Second language vocabulary learning through extensive reading: How does frequency and distribution of occurrence affect learning? *Language Teaching Research*, 19(6), 667-686. doi: 10.1177/1362168814559800.

Wesche, M. & Paribakht, T.S. 2000. Reading-Based Vocabulary Exercises in Second Language Vocabulary Acquisition. *The Modern Language Journal*, *84*(2), 196-213.

Weiss, R. E., Knowlton, D. S., & Morisson, G. R. (2002). Principles for using animation in computer-based instruction: theoretical heuristic for effective design. *Computers in Human Behavior, 18*(4), 465-477.

Widdowson, H. G. (1978). *Teaching Language as Communication*. Oxford, UK: Oxford University Press.

Wilkins, D. A. (1972). *Linguistics in Language Teaching*. London: Arnold.

Williams, M., & Burden, R. L. (1997). *Psychology for Language Teachers*. Cambridge: Cambridge University Press.

Wittrock, M. C. (1974). Learning as a Generative Process. *Educational Psychologist, 11*(2), 87-95.

Xing, P., & Fulcher, G. (2007). Reliability Assessment for Two Versions of Vocabulary Levels Tests. *System, 35*(2), 182-191.

Xu, X. (2010). The Effects of Glosses on Incidental Vocabulary Acquisition in Reading. *Journal of Language Teaching and Research, 1*(2), 117-120.

Yanguas, I. (2009). Multimedia Glosses and their Effect on L2 Text Comprehension and Vocabulary Learning. *Language Learning and Technology, 13*(2), 48-67.

Yeasmin, S., & Rahman, K. F. (2012). Triangulation Research Method as the Tool of Social Science Research. *BUP (Bangladesh University Professional) Journal, 1*(1), 154-163.

Yeh, Y., & Wang, C. (2003). Effects of Multimedia Vocabulary Annotations and Learning Styles on Vocabulary Learning. *CALICO Journal, 21*(1), 131-144.

Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *Journal of Experimental Education, 67*(3), 197-217. Retrieved online on 01 November 2016 from http://doi.org/cdtw89.

Yongqui-gu, P. (2003). Vocabulary Learning in a Second Language: Person, Task, Context and Strategies. *Teaching English as a Second or Foreign Language (TESL- EJ), 7*(2), 1-24.

Yoshii, M., & Flaitz, J. (2002). Second Language Incidental Vocabulary Retention: The effect of picture and annotation types. *CALICO Journal, 20*(1), 33-58.

Yoshii, M. (2006). L1 and L2 glosses: their effects on incidental vocabulary learning. *Language Learning and Technology, 10*(3), 85-101.

Yoshii, M. (2014). Effects of Glosses and Reviewing of Glossed Words on L2 Vocabulary Learning through Reading. *Vocabulary Learning and Instruction*, *3*(2), 19-30

Yun, J. (2011). The Effects of Hypertext Glosses on L2 Vocabulary Acquisition: a meta-analysis. *Computer Assisted Language Learning, 24*(1), 39-58.

Yusuf, M. A., Sim, T. S., & Su'ad, A. (2014). Students' Proficiency and Textual Computer Goss Use in Facilitating Vocabulary Knowledge. *English Language Teaching*. *7*(11), 99-107.

Zandieh, Z., & Jafarigohar, M. (2012). The Effects of Hypertext Gloss on Comprehension and Vocabulary Retention under Incidental and Intentional Learning Conditions. *English Language Teaching, 5*(6), 60-71.

Zarei, A., & Mahmoodzadeh, P. (2014). The Effect of multimedia glosses on L2 reading comprehension and vocabulary production. *Journal of English Language and Literature*, *1*(1), 1-7.

Zhang, J. (2004). Memory Process and the Function of Sleep. *Journal of Theoretic, 6*(6).

Zhang, B., & Li, C. (2011). Classification of L2 Vocabulary Learning Strategies: Evidence from Exploratory and Confirmatory Factor Analyses. *RELC Journal*, *42*(2), 141-154.

Zoi, M., Bellou, I., Mikropoulos, T. A. (2011). Second Language Teaching in Elementary School with a Multimedia Gloss. In D. Chen (ed.) International Conference on Languages, Literature and Linguistics IPEDR vol. 26 (pp. 54-58). Singapore: IACSIT Press.

Zwier, L. J., & Zimmerman, C. B. (2009). Bites and stings. In *Inside Reading 2: The Academic Word list in Context* (pp. 134-135). New York: Oxford University Press.

APPENDICES

Appendix A: Letter of Information & Consent Form (Pilot study)

Western Education

Project Title

Investigating the Effectiveness of Simultaneous Multimedia Glossing on L2 Learners' Vocabulary Learning and Long-term Word Retention

Letter of Information (Pilot Study)

Principal Investigator:
Farahnaz Faez, PhD, Faculty of education, University of Western Ontario
Telephone: [redacted] / E-mail: [redacted]

Study Investigator's Name: Nasrin Ramezanalialiakbar

1. **Invitation to Participate**

   You are being invited to participate in this pilot study on examining the effectiveness of simultaneous multimedia glossing on L2 learners' vocabulary learning and long-term word retention. The purpose of this pilot study is, first, to make sure that intermediate ESL (English as a second language) learners understand the English reading passages, and second, to ascertain that they have no familiarity with the underlined words in the passages.
2. **Purpose of the Letter**

   The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research.
3. **Purpose of this Study**

   The purpose of this pilot study is to examine if intermediate ESL learners understand the three reading passages. A secondary goal of this study is to make sure the participants have no familiarity with the target/unknown words in the passage.

4. **Inclusion Criteria**

   Participants who are learning English as a second language at an intermediate proficiency level, and who range between 18 to 25 years old are eligible to participate in this pilot study.

5. **Exclusion Criteria**

   Individuals who are not adult ESL learners of intermediate proficiency level will be excluded, and are not eligible to participate in the pilot study.

6. **Study Procedures**

   If you agree to participate in this pilot study, you will be asked to read the three English reading passages, and underline the words you don't know. It is anticipated that each reading passage will take 15 minutes of your time, and the entire task will take 45 minutes of your time over one session. The task(s) will be conducted in the English language center at the faculty of Education. There will be a total of 5 participants to accomplish the task.

7. **Possible Risks and Harms**

   There are no known or anticipated risks or discomforts associated with participating in this study.

8. **Possible Benefits**

   The possible benefits to participants are: (a) improving the reading comprehension skill; and (b) learning new words. The possible societal benefit may be the ability to use the learned words in social communication.

9. **Compensation**

   You will not be compensated for your participation in this research.

10. **Voluntary Participation**

    Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time with no effect on your future academic status.

11. **Confidentiality**

    All data collected will remain confidential and accessible only to the investigators of this study. If the results are published, your name will not be used. If you choose to withdraw from this study, your data will be removed and destroyed from our database.

12. **Contacts for Further Information**

    If you require any further information regarding this pilot study or your participation in the study you may contact Nasrin Ramezanali at [redacted], or call her at [redacted]. If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Research Ethics [redacted], email: [redacted]

13. **Publication**

    If the results of the pilot study are published, your name will not be used. If you would like to receive a copy of any potential study results, please contact [redacted]

Western Education

Consent Form (Pilot Study)

Project Title

Investigating the Effectiveness of Simultaneous Multimedia Glossing on L2 Learners' Vocabulary Learning and Long-term Word Retention

Study Investigator's Name: Nasrin Ramezanalialiakbar

I have read the Letter of Information, have had the nature of the study explained to me and I agree to participate. All questions have been answered to my satisfaction.

Participant's Name: ____

Participant's Signature: ____

Date: ____

Person Obtaining Informed Consent Nasrin Ramezanalialiakbar
Signature:

Date:

Appendix B: Letter of Information & Consent Form (Actual Study)

Western Education

Letter of Information (Actual Study)

Project Title:

Investigating the Effectiveness of Simultaneous Multimedia Glossing on L2 Learners' Vocabulary Learning and Long-term Word Retention

Principal Investigator:

Farahnaz Faez, PhD, Faculty of education, University of Western Ontario

Telephone: [redacted] / E-mail: [redacted]

Study Investigator's Name: Nasrin Ramezanalialiakbar

<u>Letter of Information</u>

1. **Invitation to Participate**
   You are being invited to participate in this PhD study on examining the effectiveness of simultaneous multimedia glossing on L2 learners' vocabulary learning and long-term word retention.

2. **Purpose of the Letter**
   The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research.

3. **Purpose of this Study**
   The purpose of this study is to examine the effectiveness of meaningful, contextual and multimodal learning tools of simultaneous textual, aural, and video/animation glossing to improve L2 vocabulary learning. A secondary goal of this study is to investigate whether simultaneous multimedia glossing fosters L2 learners' short and long-term word retention when the instructional sessions are distributed in time.

4. **Inclusion Criteria**
   Participants who are learning English as a second/foreign language (ESL/EFL) at an intermediate proficiency level, and who range between 18 to 25 years old are eligible to participate in this study.

5. Exclusion Criteria

Individuals who are not adult ESL/ EFL learners of intermediate proficiency level will be excluded, and thus are not eligible to participate in this study.

6. Study Procedures

If you agree to participate in this four-week study,

1) You will first be assigned to 4 groups of A (control group), and B, C and D (experimental groups), and be given name tags to wear on the days of instruction.

2) You will then be asked to fill out a demographic information form including your age, gender, length of time studying English, and how many other languages you know in addition to English. The estimated time to complete the form is 3-5 minutes.

3) You will be given a vocabulary levels test (VLT) to fill. The test includes 30 word items. The approximate time to complete the test will be between 25-30 minutes.

4) After completing the vocabulary levels test, you will be given two other vocabulary pre-tests (productive recall and MC productive recognition tests). Each test consists of approximately 15-20 question items, and the estimated time to complete each test will be 15-20 minutes.

5) I will instruct the new words in one week with three consecutive sessions, every other day. Each instructional session will take approximately 30 minutes, and the total length of the instruction would be 90 minutes.

6) After the instructions on each day, I will give you two vocabulary post-tests (recall productive test and MC productive recognition test) from the same text that day. Each test will take approximately 15-20 minutes for the participants to complete. At the end of the third instructional session, I will ask the participants to meet again two weeks (14 days) later.

7) Two weeks after the instruction (week 3), on a scheduled day, you will be asked to do two post-tests (recall productive test and MC productive recognition test). Each test will take approximately 15-20 minutes for your time to complete.

8) After the delayed post-tests, on the same day, a questionnaire will be distributed to you to fill out. It will take approximately 10-15 minutes for them to complete it.

9) I will also ask you to indicate your preference and availability for an interview for the following week (week 4). The interview will take 10-15 minutes of your time. The interview will be audio-recorded. You can also take part in interview session if you do not wish to be audio recorded.

7. Possible Risks and Harms

There are no known or anticipated risks or discomforts associated with participating in this study.

8. Possible Benefits

The possible benefits to participants of this study are: (a) improving the reading comprehension skill; (b) learning new words; (c) experiencing new vocabulary instructional modes; and (d) enhancing long-term word retention. The possible societal benefit may be the ability to use the learned words in social communication, and retrieve them whenever required easily and fast.

9. Compensation

You will be given a small gift for your participation and time in this research.

10. Voluntary Participation

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time with no effect on your future academic status. Besides, if you do not want to be audio-recorded, you may not participate in this study.

11. Confidentiality

All data collected will remain confidential and accessible only to the investigators of this study. If the results are published, your name will not be used. If you choose to withdraw from this study, your data will be removed and destroyed from our database.

12. Contacts for Further Information

If you require any further information regarding this study or your participation in the study you may contact Nasrin Ramezanali at [redacted] or call her at [redacted]. If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Research Ethics [redacted] email: [redacted]

13. Publication

If the results of this study are published, your name will not be used. If you would like to receive a copy of any potential study results, please contact [redacted]

Western Education

Consent Form

Project Title:

Investigating the Effectiveness of Simultaneous Multimedia Glossing on L2 Learners' Vocabulary Learning and Long-term Word Retention

Study Investigator's Name: Nasrin Ramezanalialiakbar

I have read the Letter of Information, have had the nature of the study explained to me and I agree to participate. All questions have been answered to my satisfaction.

Participant's Name: ____________________________

Participant's Signature: ____________________________

Date: ____________________________

Person Obtaining Informed Consent Nasrin Ramezanalialiakbar

Signature:

Date: 07/ /2015

Appendix C: Demographic Information Form

Demographic Information Form

Please respond to the following questions briefly. Estimated time: 5-7 minutes

1. How old are you? -----------------

2. What is your gender? M…… F……

3. How long have you studied English? ----------------------

4. How many other languages do you know in addition to English? ………………

5. Have you visited/lived/or studied in English speaking countries?

Appendix D: Reading Text 1 (The Silk Road)

The Silk Road

In the **ruins** of the ancient Roman city of Pompeii, which was destroyed by a **volcano** in the year 79 C.E., a mirror was found. It had an **ivory** handle in the shape of a female **goddess**. The mirror was from India. In the **tomb** of Li Xian, a Chinese military official who died in 569 C.E., **archeologists** found a water **pitcher** in the shape of a **vase**. The pitcher had a combination of different styles: the shape was from Persia (today's Iran), many details were from central Asia, and the figures on the side were Greek stories from the Trojan War. In the Japanese city of Nara, the 8th century Shosoin Treasures households thousands of exquisite objects of great beauty- furniture, musical instruments, **weapons**, fabric, and military **armor**. These objects come from what is today Vietnam, Western China, Iraq, the Roman Empire, and Egypt. Clearly, long before the globalization of our modern world, trade was going on between very distant lands, and the objects tell a story about a place and time.

From ancient times, cultures have influenced each other along the famous Silk Road, although it was not truly one continuous road. Instead, it was a 5,000-mile series or network of trails that connected East Asia to the Mediterranean. In ancient times, it was never called the "Silk Road". The term *Silk Road* was coined in the 19th century by a German explorer. He was thinking of one of the goods that people in the west found especially desirable- silk fabric from China. For centuries, the Chinese kept as a secret the way in which silk is produced. They exchanged this fabric in Mediterranean glass, whose production was also kept secret by the Romans. However, **merchants** also moved many other goods along these trade routes: spices (such as cinnamon), musical instruments, tea, valuable stones, wool, **linen**, and other fabrics. Ideas and knowledge also moved along the Silk Road. Travelers to foreign regions took with them ideas about art, architecture, styles of living and religion.

Source

Hartman, P., & Kirn, E. (2014). Interactions Reading (2). pp. 171-172. The McGraw-Hills companies.

Appendix E: Reading Text 2 (Problems in the natural world)

Problems in the natural world

Honeybees are the most important **pollinators** in most regions of the world where flowering plants exist. However, they are by no means the only insects that play this role. Flies, butterflies, **beetles**, **wasp**, **bumblebees**, and even ants can also pollinate plants. Very few flowers are dependent on a single insect species, although no other pollinators are as effective as are honeybees. In all, 80% of flowering plants worldwide are pollinated by insects and of these about 85% by honeybees. As many as 90% of fruit-trees and flowers are dependent on honeybees. The list of flowering plant pollinated by honeybees includes 170,000 species. The number of flowering plant species that are dependent on honeybees, and without which they would do badly, is estimated to be about 40,000. This worldwide sea of flowers is pollinated by just nine species and in Europe and Africa by only one, which is indispensable for most flowering plants. The fact that honeybees are so successful at pollinating means there is little room for **competitors** wanting to do the same job.

The absence of honeybees from an ecosystem can have an extremely negative impact on human beings. A clear example can be found in southern Sichuan in China. Every year in April, thousands of people take **feather**, **dusters** and **ladders** into the pear **orchards** and climb the trees. They use the dusters to brush each individual tree in order to collect pollen that will be dried and transferred to other trees. It is a slow and boring job that is normally done by honeybees. More than 20 years ago, **pesticides** killed all the honeybees of Sichuan. Problems with honeybee populations are occurring all around the world. The US has lost at least 35% of its honeybees in recent years. Canada, Brazil, India and China have also lost huge numbers of bees, as has Western Europe. In France, losses of up to 60% have been estimated, while in the UK the government has said bees could completely disappear in less than ten years' time. In recent years, environmentalists have focused on greenhouse gases and the warming planet, making them less aware of the issues surrounding bees and pollination. Klein (2007) has confirmed the seriousness of failing bee population. She found that three quarters of the world's 115 most important **crops** require animal pollination and that bees are the most useful pollinators of commercial crops around the world.

Source:

Thaine, C. (2012). Cambridge Academic English an integrated skills course for EAP, Excerpt taken from pp. 30-33. Cambridge University Press.

Appendix F: Reading text 3 (Bites and Stings)

Bites and Stings

The young woman had been looking forward to her nice new apartment in Manhattan. Circumstances turned out to be less comfortable than she expected, as this posting to an online forum about insect bites shows.

I just moved into a newly renovated apartment and got 10 huge, itchy bug bites on my arms, legs, and ***hip****. I thought it was my mattress, so I got rid of it and bought a new one. I tried freezing out my apartment by leaving the door open during the winter chill since I heard the bugs can't survive in temps less than 25 degrees. I went to* ***a dermatologist*** *who said the bite pattern isn't like any of the usual apartment pests, and he didn't know what it was. My immune system has reacted to the bites, and I have prickly itching all over my body. If anyone has found the solution, please email me. Thank you!*

Throughout North America, countless people crawl into bed at night knowing exactly how the writer feels. Instead of a peaceful night's sleep, they will get a new round of bites by some mysterious pest. Bites by insects or **arachnids** such as **fleas**, ticks, **horseflies**, **mosquitos**, or **bedbugs** are extremely common. A bite, which involves a creature's mouth parts, is different from sting, which is made with a sharp structure appended to a creature's rear end. Most insect bites cause discomfort, if any at all. The bite might cause a little swelling because chemicals in the bug's saliva irritate the skin. When they do happen, the bite victim's own behavior might be to blame. Insects and arachnids account for almost all the bites North Americans suffer, but they are identified less easily than any others. If a dog, a **rodent**, a horse, or even a snake bites you, you know that it has happened and which creature did it. If an insect or spider bites you, you may not feel the contact of its mouth parts with your skin. It is often difficult even to tell whether the bite was an insect. For example, consider the case of the brown recluse spider and the **deer** tick. The brown spider is able to inject enough powerful **venom** to cause serious medical problems for a healthy adult. The deer tick can carry the bacterium responsible for Lyme disease.

Source

Zwier, L., J & Zimmerman, C., B. (2009). Bites and Stings. *Inside Reading 2: The Academic Word list in Context* (pp. 134-135). Oxford University Press.

Appendix G: Vocabulary Levels Test (VLT)

Vocabulary Levels Test

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning.

Here is an example:

1. Business	______________	part of a house
2. Clock	______________	animal with
3. Horse		four legs
4. Pencil	______________	something used
5. Shoe		for writing
6. Wall		

You answer in the following way:

1. Business	______6______	part of a house
2. Clock	______3______	animal with
3. Horse		four legs
4. Pencil	______4______	something used
5. Shoe		for writing
6. Wall		

Now start answering the questions below. Put a number from the left on the blanks.

1. Copy	_______	end or highest
2. Event		point
3. Motor	_______	this moves a
4. Pity		car
5. Profit	_______	thing made to
6. Tip		be like another

1. Accident	_______	loud deep
2. Debt		sound
3. Fortune	_______	something
4. Pride		you must pay
5. Roar	_______	having a high
6. Thread		opinion of
		yourself

1. Coffee
2. Disease
3. Justice
4. Skirt
5. Stage
6. Wage

______ money for work
______ a piece of clothing
______ using the law in the right way

1. Arrange
2. Develop
3. Lean
4. Owe
5. Prefer
6. Seize

______ grow
______ put in order
______ like more than something else

1. Clerk
2. Frame
3. Noise
4. Respect
5. Theater
6. wine

______ a drink
______ office worker
______ unwanted sound

1. Blame
2. Elect
3. Jump
4. Threaten
5. Melt
6. Manufacture

______ make
______ choose by voting
______ become like water

1. Dozen
2. Empire
3. Gift
4. Tax
5. Relief
6. Opportunity

______ chance
______ twelve
______ money paid to the government

1. Ancient
2. Curious
3. Difficult
4. Entire
5. Holy
6. Social

______ not easy
______ very old
______ related to God

1. Admire
2. Complain
3. Fix
4. Hire
5. Introduce
6. Stretch

______ make wider or longer
______ bring in for the first time
______ have a high opinion of someone

1. Slight
2. Bitter
3. Lovely
4. Merry
5. Popular
6. independent

______ beautiful
______ small
______ like by many people

Appendix H: Pre-test Productive Recall Vocabulary Test

Productive Recall Vocabulary Test

Estimated time: 15-20 minutes

Instruction: Please read the definitions, and write the word in the space provided. Each single space represents one letter of the word.

1. A small insect with a hard covering on its back

   _ _ _ _ _ _

2. A soft part that covers a bird's body

   _ _ _ _ _ _ _

3. An insect that bites animals such as cow and sheep

   _ _ _ _ _ _ _ _

4. A doctor who studies and treat skin diseases

   _ _ _ _ _ _ _ _ _ _ _ _ _

5. A plant that is grown as food, such as a fruit or a vegetable

   _ _ _ _

6. A woman who is loved, especially for her beauty

   _ _ _ _ _ _ _

7. A person or an organization that takes part in a match

   _ _ _ _ _ _ _ _ _ _

8. The connection at the top of the leg

   _ _ _

9. A class of insects that include spiders

   _ _ _ _ _ _ _ _

10. A cloth for removing spot, oil and dirt

    _ _ _ _ _ _

11. A piece of land or a garden where fruits are grown

    _ _ _ _ _ _ _

12. A tool for climbing up and down a wall or a tree

_ _ _ _ _ _

13. Remains of a destroyed building or a town

_ _ _ _ _

14. A poisonous liquid produced by some snakes and spiders when they bite

_ _ _ _ _

15. A bee that causes plants to make fruit or seed

_ _ _ _ _ _ _ _ _ _

16. An object used for fighting such as a knife, gun or bomb

_ _ _ _ _ _

17. A person who buys and sells products

_ _ _ _ _ _ _ _

18. A mountain with a large opening which sends out burned materials

_ _ _ _ _ _ _

19. A creamy white color material that makes the tooth of an elephant

_ _ _ _ _

20. A container with a handle for holding water

_ _ _ _ _ _ _

21. An insect with wings that causes diseases such as Malaria and yellow fever

_ _ _ _ _ _ _ _

22. A container for holding flowers

_ _ _ _

23. A material for killing insects

_ _ _ _ _ _ _ _ _

24. A large stone under which someone is buried

_ _ _ _

25. A black and yellow flying insect _ _ _ _

26. A type of small animal with sharp front teeth such as a mouse

— — — — — —

27. A type of cloth or sheet made from fiber

— — — — —

28. An insect without wings that feeds on animals

— — — —

29. An animal with long legs that eats grass, and runs fast

— — — —

30. A large hairy flying insect that does not bite

— — — — — — — — —

31. A small insect that feeds on people's blood

— — — — — —

32. A person who studies buildings, animals and objects of the past

— — — — — — — — — — — —

33. Metal covering worn by soldiers to protect the body

— — — — —

Good Luck

Score ____

Appendix I: Pre-test Multiple-choice Productive Recognition Test

Multiple-choice Productive Recognition Vocabulary Test

Estimated time: 15-20 minutes

Instruction: For each question, read the definition, and choose the word that best matches with the definition.

1. The connection at the top of the leg
   a. Rib b. Hip c. Knee d. Ankle

2. A woman who is loved, especially for her beauty
   a. Fairy b. Idol c. Goddess d. Angel

3. A person who buys and sells products
   a. Shepherd b. Merchant c. Clerk d. Shopkeeper

4. A material used for killing insects
   a. Trap b. Medicine c. Pesticide d. Fertilizer

5. A plant that is grown as food, such as a fruit or a vegetable
   a. Herb b. Seed c. Harvest d. Crop

6. A person or an organization that takes part in a match
   a. Warrior b. Solider c. Competitor d. Administrator

7. A bee that causes plants to make fruit or seed
   a. Pollinator b. Worker bee c. Honey bee d. Bumble bee

8. A type of small animal with sharp front teeth such as a mouse
   a. Squirrel b. Rodent c. Rabbit d. Raccoon

9. Remains of a destroyed building or town
   a. Stones b. Ruins c. Bricks d. Pieces

10. A poisonous liquid produced by some snakes and spiders when they bite
    a. Drug b. Venom c. Salvia d. Toxin

11. A black and yellow flying insect
    a. Bee b. Wasp c. Beetle d. Fly

12. A tool for climbing up and down a wall or a tree
    a. Elevator b. Stair c. Ladder d. Escalator

13. A large hairy flying insect that does not bite
    a. Bumble bee b. Honey bee c. Carpenter bee d. Wild bee

14. An animal with long legs that eats grass, and runs fast
    a. Goat b. Giraffe c. Deer d. Rodent

15. A person who studies buildings, animals and objects of the past
    a. Archeologist b. Anthropologist c. Entomologist d. Scientist

16. A class of insects that include spiders
    a. Mosquito b. Tick c. Beetle d. Arachnid

17. A metal covering worn by soldiers to protect the body
    a. Shield b. Armor c. Helmet d. Gown

18. A cloth for removing spot, soil and dirt
    a. Sponge b. Duster c. Broom d. Wiper

19. A doctor who studies and treats skin diseases
    a. Beautician b. Specialist c. Dermatologist d. Cosmetic surgeon

20. An insect without wings that feeds on animals
    a. Flea b. Worm c. Bug d. Mosquito

21. A mountain with a large opening which sends out burned materials
    a. Wildfire b. Volcano c. Hurricane d. Thunderstorm

22. An object used for fighting such as a knife, gun or bomb
    a. Weapon b. Rifle c. Armor d. Furniture

23. A large stone under which someone is buried
    a. Hole b. Rock c. Cemetery d. Tomb

24. A piece of land or a garden where fruits are grown
    a. Orchard b. Pasture c. Field d. Ground

25. A long-legged insect with wings that cause diseases such as Malaria and yellow fever
    a. Flea b. Wasp c. Mosquito d. Beetle

26. A type of cloth or sheet made from fiber
    a. Spice b. Linen c. Stone d. Wood

27. A container with a handle for holding water
    a. Pitcher b. Vase c. Ladder d. Pesticide

28. A small insect with a hard covering on its back
    a. Wasp b. Flea c. Beetle d. Bedbug

29. A soft part that covers a bird's body
    a. Fur b. Feather c. Duster d. Pitcher

30. A creamy white color material that makes the tooth of an elephant
    a. Ivory b. Linen c. Vase d. Cinnamon

31. A container for holding flowers
    a. Venom b. Vase c. Vessel d. Crop

32. A small insect that feeds on people's blood
    a. Bumble bee b. Lady bug c. Bedbug d. Water bug

33. An insect that bites animals such as cow and sheep
    a. Tick b. Fruit fly c. Beetle d. Horse fly

Good Luck

Score ____

Appendix J: Text 1 (Productive Recall Vocabulary Test)

Productive Recall Vocabulary Test

Estimated time: 15-20 minutes

Instruction: Please read the definitions, and write the word in the space provided. Each single space represents one letter of the word.

1. A type of cloth or sheet made from fiber

   — — — — —

2. Metal covering worn by soldiers to protect the body

   — — — — —

3. An object used for fighting such as a knife, gun or bomb

   — — — — — —

4. A person who buys and sells products

   — — — — — — — —

5. A creamy white color material that makes the tooth of an elephant

   — — — — —

6. A mountain with a large opening which sends out burned materials

   — — — — — — —

7. Remains of a destroyed building or a town

   — — — — —

8. A woman who is loved, especially for her beauty

   — — — — — — —

9. A container with a handle for holding water

   — — — — — — —

10. A person who studies buildings, animals and objects of the past

    — — — — — — — — — — — —

11. A large stone under which someone is buried

    — — — —

12. A container for holding flowers

    — — — —

Good Luck

Score ____

Appendix J: Text 1 (Multiple-choice Productive Recognition Vocabulary Test)

Multiple-choice Productive Recognition Vocabulary Test

Estimated time: 15-20 minutes

Instruction: For each question, read the definition, and choose the word that best matches with the definition.

1. A woman who is loved, especially for her beauty
   a. Fairy b. Idol c. Goddess d. Angel

2. A person who buys and sells products
   a. Shepherd b. Merchant c. Clerk d. Shopkeeper

3. Remains of a destroyed building or town
   a. Stones b. Ruins c. Bricks d. Pieces

4. A person who studies buildings, animals and objects of the past
   a. Archeologist b. Anthropologist c. Entomologist d. Scientist

5. A mountain with a large opening which sends out burned materials
   a. Wildfire b. Volcano c. Hurricane d. Thunderstorm

6. A large stone under which someone is buried
   a. Hole b. Rock c. Cemetery d. Tomb

7. A type of cloth or sheet made from fiber
   a. Spice b. Linen c. Stone d. Wood

8. A container with a handle for holding water
   a. Pitcher b. Vase c. Ladder d. Pesticide

9. A creamy white color material that makes the tooth of an elephant
   a. Ivory b. Linen c. Vase d. Cinnamon

10. An object used for fighting such as a knife, gun or bomb
    a. Weapon b. Rifle c. Armor d. Furniture

11. A container for holding flowers
    a. Venom b. Vase c. Vessel d. Crop

12. A metal covering worn by soldiers to protect the body
    a. Shield b. Armor c. Helmet d. Gown

Good Luck

Score ____

Appendix K: Text 2 (Productive Recall Vocabulary Test)

Productive Recall Vocabulary Test

Estimated time: 15-20 minutes

Instruction: Please read the definitions, and write the word in the space provided. Each single space represents one letter of the word.

1. A small insect with a hard covering on its back

   _ _ _ _ _ _

2. A soft part that covers a bird's body

   _ _ _ _ _ _ _

3. A plant that is grown as food, such as a fruit or a vegetable

   _ _ _ _

4. A person or an organization that takes part in a match

   _ _ _ _ _ _ _ _ _ _

5. A cloth for removing spot, oil and dirt

   _ _ _ _ _ _

6. A tool for climbing up and down a wall or a tree

   _ _ _ _ _ _

7. A piece of land or a garden where fruits are grown

   _ _ _ _ _ _ _

8. A bee that causes plants to make fruit or seed

   _ _ _ _ _ _ _ _ _ _

9. A black and yellow flying insect

   _ _ _ _

10. A material for killing insects

    _ _ _ _ _ _ _ _ _

11. A large hairy flying insect that does not bite

    _ _ _ _ _ _ _ _ _

Good Luck

Score ____

Appendix K: Text 2 (Multiple-choice Productive Recognition Vocabulary Test)

Multiple-choice Productive Recognition Vocabulary Test

Estimated time: 15-20 minutes

Instruction: For each question, read the definition, and choose the word that best matches with the definition.

1. A material used for killing insects
   a. Trap b. Medicine c. Pesticide d. Fertilizer
2. A person or an organization that takes part in a match
   a. Warrior b. Solider c. Competitor d. Administrator
3. A plant that is grown as food, such as a fruit or a vegetable
   a. Herb b. Seed c. Harvest d. Crop
4. A bee that causes plants to make fruit or seed
   a. Pollinator b. Worker bee c. Honey bee d. Bumble bee
5. A black and yellow flying insect
   a. Bee b. Wasp c. Beetle d. Fly
6. A tool for climbing up and down a wall or a tree
   a. Elevator b. Stair c. Ladder d. Escalator
7. A large hairy flying insect that does not bite
   a. Bumble bee b. Honey bee c. Carpenter bee d. Wild bee
8. A cloth for removing spot, soil and dirt
   a. Sponge b. Duster c. Broom d. Wiper
9. A piece of land or a garden where fruits are grown
   a. Orchard b. Pasture c. Field d. Ground
10. A soft part that covers a bird's body
    a. Fur b. Feather c. Duster d. Pitcher
11. A small insect with a hard covering on its back
    a. Wasp b. Flea c. Beetle d. Bedbug

Good Luck

Score ____

Appendix L: Text 3 (Productive Recall Vocabulary Test)

Multiple-choice Productive Recognition Vocabulary Test

Estimated time: 15-20 minutes

Instruction: For each question, read the definition, and choose the word that best matches with the definition.

1. A long-legged insect with wings that cause diseases such as Malaria and yellow fever

   a. Flea b. Wasp c. Mosquito d. Beetle

2. A doctor who studies and treats skin diseases

   a. Beautician b. Specialist c. Dermatologist d. Cosmetic surgeon

3. An insect without wings that feeds on animals

   a. Flea b. Worm c. Bug d. Mosquito

4. An animal with long legs that eats grass, and runs fast

   a. Goat b. Giraffe c. Deer d. Rodent

5. A type of small animal with sharp front teeth such as a mouse

   a. Squirrel b. Rodent c. Rabbit d. Raccoon

6. The connection at the top of the leg

   a. Rib b. Hip c. Knee d. Ankle

7. A class of insects that include spiders

   a. Mosquito b. Tick c. Beetle d. Arachnid

8. A small insect that feeds on people's blood

   a. Bumble bee b. Lady bug c. Bedbug d. Water bug

9. A poisonous liquid produced by some snakes and spiders when they bite

   a. Drug b. Venom c. Salvia d. Toxin

10. An insect that bites animals such as cow and sheep

    a. Tick b. Fruit fly c. Beetle d. Horse fly

Good Luck

Score ____

Appendix L: Text 3 (Multiple-choice Productive Recognition Vocabulary Test)

Multiple-choice Productive Recognition Vocabulary Test

Estimated time: 15-20 minutes

Instruction: For each question, read the definition, and choose the word that best matches with the definition.

1. A long-legged insect with wings that cause diseases such as Malaria and yellow fever
   a. Flea b. Wasp c. Mosquito d. Beetle
2. A doctor who studies and treats skin diseases
   a. Beautician b. Specialist c. Dermatologist d. Cosmetic surgeon
3. An insect without wings that feeds on animals
   a. Flea b. Worm c. Bug d. Mosquito
4. An animal with long legs that eats grass, and runs fast
   a. Goat b. Giraffe c. Deer d. Rodent
5. A type of small animal with sharp front teeth such as a mouse
   a. Squirrel b. Rodent c. Rabbit d. Raccoon
6. The connection at the top of the leg
   a. Rib b. Hip c. Knee d. Ankle
7. A class of insects that include spiders
   a. Mosquito b. Tick c. Beetle d. Arachnid
8. A small insect that feeds on people's blood
   a. Bumble bee b. Lady bug c. Bedbug d. Water bug
9. A poisonous liquid produced by some snakes and spiders when they bite
   a. Drug b. Venom c. Salvia d. Toxin
10. An insect that bites animals such as cow and sheep
    a. Tick b. Fruit fly c. Beetle d. Horse fly

Good Luck

Score ____

Appendix M: Post-test Productive Recall Vocabulary Test

Productive Recall Vocabulary Test

Estimated time: 15-20 minutes

Instruction: Please read the definitions, and write the word in the space provided. Each single space represents one letter of the word.

1. A person or an organization that takes part in a match
   — — — — — — — — — —
2. A piece of land or a garden where fruits are grown
   — — — — — — —
3. The connection at the top of the leg
   — — —
4. A doctor who studies and treat skin diseases
   — — — — — — — — — — — — —
5. A small insect with a hard covering on its back
   — — — — — —
6. A soft part that covers a bird's body
   — — — — — — —
7. A small insect that feeds on people's blood
   — — — — — —
8. A person who studies buildings, animals and objects of the past
   — — — — — — — — — — — —
9. Metal covering worn by soldiers to protect the body
   — — — — —
10. A person who buys and sells products
    — — — — — — — —
11. A mountain with a large opening which sends out burned materials
    — — — — — — —
12. A bee that causes plants to make fruit or seed
    — — — — — — — — — —
13. A type of cloth or sheet made from fiber
    — — — — —

14. An insect without wings that feeds on animals

_ _ _ _

15. An insect that bites animals such as cow and sheep

_ _ _ _ _ _ _ _

16. A plant that is grown as food, such as a fruit or a vegetable

_ _ _ _

17. A woman who is loved, especially for her beauty

_ _ _ _ _ _ _

18. Remains of a destroyed building or a town

_ _ _ _ _

19. A poisonous liquid produced by some snakes and spiders when they bite

_ _ _ _ _

20. A cloth for removing spot, oil and dirt

_ _ _ _ _ _

21. A tool for climbing up and down a wall or a tree

_ _ _ _ _ _

22. A class of insects that include spiders

_ _ _ _ _ _ _ _

23. An animal with long legs that eats grass, and runs fast

_ _ _ _

24. A large hairy flying insect that does not bite

_ _ _ _ _ _ _ _ _

25. An object used for fighting such as a knife, gun or bomb

_ _ _ _ _ _

26. A material for killing insects

_ _ _ _ _ _ _ _ _

27. A large stone under which someone is buried

_ _ _ _

28. A black and yellow flying insect

_ _ _ _

29. A container for holding flowers

_ _ _ _

30. A creamy white color material that makes the tooth of an elephant

_ _ _ _ _

31. A container with a handle for holding water

_ _ _ _ _ _ _

32. An insect with wings that causes diseases such as Malaria and yellow fever

_ _ _ _ _ _ _ _

33. A type of small animal with sharp front teeth such as a mouse

_ _ _ _ _ _

Good Luck

Score ____

Appendix N: Post-test Multiple-choice Productive Recognition Vocabulary Test

Multiple-choice Productive Recognition Vocabulary Test

Estimated time: 15-20 minutes

Instruction: For each question, read the definition, and choose the word that best matches with the definition.

1. A type of small animal with sharp front teeth such as a mouse
   a. Squirrel b. Rodent c. Rabbit d. Raccoon

2. A black and yellow flying insect
   a. Bee b. Wasp c. Beetle d. Fly

3. A material used for killing insects
   a. Trap b. Medicine c. Pesticide d. Fertilizer

4. A bee that causes plants to make fruit or seed
   a. Pollinator b. Worker bee c. Honey bee d. Bumble bee

5. A tool for climbing up and down a wall or a tree
   a. Elevator b. Stair c. Ladder d. Escalator

6. A person who studies buildings, animals and objects of the past
   a. Archeologist b. Anthropologist c. Entomologist d. Scientist

7. A cloth for removing spot, soil and dirt
   a. Sponge b. Duster c. Broom d. Wiper

8. An object used for fighting such as a knife, gun or bomb
   a. Weapon b. Rifle c. Armor d. Furniture

9. A piece of land or a garden where fruits are grown
   a. Orchard b. Pasture c. Field d. Ground

10. A soft part that covers a bird's body
    a. Fur b. Feather c. Duster d. Pitcher

11. A creamy white color material that makes the tooth of an elephant
    a. Ivory b. Linen c. Vase d. Cinnamon

12. A container for holding flowers
    a. Venom b. Vase c. Vessel d. Crop

13. A small insect that feeds on people's blood
    a. Bumble bee b. Lady bug c. Bedbug d. Water bug

14. A long-legged insect with wings that cause diseases such as Malaria and yellow fever
    a. Flea b. Wasp c. Mosquito d. Beetle

15. The connection at the top of the leg
    a. Rib b. Hip c. Knee d. Ankle

16. A woman who is loved, especially for her beauty
    a. Fairy b. Idol c. Goddess d. Angel

17. A person who buys and sells products
    a. Shepherd b. Merchant c. Clerk d. Shopkeeper

18. Remains of a destroyed building or town
    a. Stones b. Ruins c. Bricks d. Pieces

19. A large hairy flying insect that does not bite
    a. Bumble bee b. Honey bee c. Carpenter bee d. Wild bee

20. A poisonous liquid produced by some snakes and spiders when they bite
    a. Drug b. Venom c. Salvia d. Toxin

21. A plant that is grown as food, such as a fruit or a vegetable
    a. Herb b. Seed c. Harvest d. Crop

22. A person or an organization that takes part in a match
    a. Warrior b. Solider c. Competitor d. Administrator

23. An animal with long legs that eats grass, and runs fast
    a. Goat b. Giraffe c. Deer d. Rodent

24. An insect without wings that feeds on animals
    a. Flea b. Worm c. Bug d. Mosquito

25. A mountain with a large opening which sends out burned materials

a. Wildfire b. Volcano c. Hurricane d. Thunderstorm

26. A type of cloth or sheet made from fiber

a. Spice b. Linen c. Stone d. Wood

27. A class of insects that include spiders

a. Mosquito b. Tick c. Beetle d. Arachnid

28. A metal covering worn by soldiers to protect the body

a. Shield b. Armor c. Helmet d. Gown

29. A doctor who studies and treats skin diseases

a. Beautician b. Specialist c. Dermatologist d. Cosmetic surgeon

30. A large stone under which someone is buried

a. Hole b. Rock c. Cemetery d. Tomb

31. An insect that bites animals such as cow and sheep

a. Tick b. Fruit fly c. Beetle d. Horse fly

32. A container with a handle for holding water

a. Pitcher b. Vase c. Ladder d. Pesticide

33. A small insect with a hard covering on its back

a. Wasp b. Flea c. Beetle d. Bedbug

Good Luck

Score ____

Appendix O: List of Words & Definitions

Word-List Definitions[73]

1. **Archeologist**: a person who studies buildings, animals and objects of the past
2. **Arachnid**: a class of insects that include spiders
3. **Armor**: metal-covering worn by soldiers to protect the body
4. **Bedbug**: a small insect that feeds on people's blood
5. **Beetle**: a small insect with a hard covering on its back
6. **Bumblebee**: a large hairy flying insect that does not bite
7. **Crop**: a plant that is grown as food, such as a fruit or a vegetable
8. **Competitor**: a person or an organization that takes part in a match
9. **Deer**: an animal with long legs that eats grass, and runs fast
10. **Dermatologist**: a doctor who studies and treats skin diseases
11. **Duster**: a cloth for removing spot, soil and dirt
12. **Feather**: a soft part that covers a bird's body
13. **Flea**: an insect without wings that feeds on the blood of animals
14. **Goddess**: a woman who is loved, especially for her beauty
15. **Hip**: the connection at the top of the leg
16. **Horsefly**: an insect that bites animals such as cows and sheep
17. **Ivory**: a creamy white color material that makes the tooth of an elephant
18. **Ladder**: a tool for climbing up and down a wall or a tree
19. **Linen**: a type of cloth or sheet made from fiber
20. **Merchant**: a person who buys and sells products
21. **Mosquito**: an insect with wings that causes diseases such as Malaria and yellow fever
22. **Orchard**: a piece of land or a garden where fruits are grown
23. **Pesticide**: a material used for killing insects

---

[73] All the target words are in alphabetical order.

24. **Pitcher**: a container with a handle for holding water
25. **Pollinator**: a bee that causes plants to make fruit or seed
26. **Rodent**: a type of small animal with sharp front teeth such as a mouse
27. **Ruins**: remains of a destroyed building or town
28. **Tomb**: a large stone under which someone is buried
29. **Vase**: a container for holding flowers
30. **Venom**: a poisonous liquid produced by some snakes and spiders when they bite
31. **Volcano**: a mountain with a large opening which sends out burned materials
32. **Wasp**: a black and yellow flying insect
33. **Weapon**: an object used for fighting such as a knife, gun, or bomb

Appendix P : Questionnaire

Questionnaire

Purpose of the questionnaire

You are invited to fill out this questionnaire. The purpose of this questionnaire is to gain a better understanding of your attitudes and preferences towards the three modes of vocabulary instruction (text-definition alone, text-definition and the audio pronunciation of the word, and text-definition and video/animation of the word). Participation in this study is voluntary. You may refuse to participate, or answer any questions or withdraw from the study at any time with no effect on your academic status.

Confidentiality

All data collected will remain confidential and accessible only to the investigators of this study. If the results are published, your name will not be used. If you choose to withdraw from this study, your data will be removed and destroyed from our database.

Contacts for Further Information

If you require any further information regarding this questionnaire or your participation in the study, you may contact Nasrin Ramezanali at [redacted]. The questionnaire will take 10 to 15 minutes of your time.

A. For questions 1 to 9 below, please state the extent to which you agree or disagree with the following statements using the scale of 1 to 5 where:

1= strongly agree; 2= agree; 3 = neither agree nor disagree; 4 = disagree; 5 = strongly disagree

	strongly agree	**agree**	**neither agree nor disagree**	**disagree**	**strongly disagree**
1. It is easy for me to learn new words with text-definition alone.					
2. It is easy for me to learn new words with text-definition and audio pronunciation.					
3. It is easy for me to learn new words with text-definition and video/animation.					
4. It was easier for me to remember words in the final test when instructed by text-definition alone.					
5. It was easier for me to remember words in the final test when instructed by text-definition and audio pronunciation.					
6. It was easier for me to remember words in the final test when instructed by text- definition and video/animation.					
7. Given the choice, I would use text- definition alone technique to learn new words in the future.					
8. Given the choice, I would use text- definition and audio pronunciation technique to learn new words in the future.					
9. Given the choice, I would use text- definition and video/animation technique to learn new words in the future.					

B. For questions 10, 11, and 12, please state the extent to which you found the vocabulary learning helpful using the scale of 1 to 5 where:

1= extremely helpful; 2=helpful; 3= somewhat helpful; 4= neither helpful nor unhelpful; 5=unhelpful

10. How helpful was definition alone in learning new words?

1. 2. 3. 4. 5.

11. How helpful was definition with audio pronunciation in learning new words?

1. 2. 3. 4. 5.

12. How helpful was definition with video/animation in learning new words?

1. 2. 3. 4. 5.

C. For question 13, of the three choices listed below, please indicate your first, second and third choice for learning new words by placing the numbers 1, 2, and 3 next to each choice:

13. A. Text- definition alone

    B. Text- definition with audio pronunciation of the new words

    C. Text- definition with video/animation of the new words

14. Please state the reason for your selection. Which mode of vocabulary instruction helped you learn and remember words easier? Why?

15. Are there any other comments about the modes of vocabulary learning that you would like to add?

16. Please mention your interest and availability for a face-to-face interview for the following week.

Thank you for your cooperation.

Appendix Q: Face-to-face Semi-structured Interview Guide

Face-to-face Semi-Structured Interview Guide

Purpose of the interview

You are invited to participate in this interview. The purpose is to seek your perceptions on what you perceived to be the most useful mode of vocabulary instruction. Participation in this interview is voluntary. You may refuse to participate, or answer any questions or withdraw from the interview at any time with no effect on your future academic status.

Confidentiality

All data collected will remain confidential and accessible only to the investigators of this study. If the results are published, your name will not be used. If you choose to withdraw from this study, your data will be removed and destroyed from our database.

Contacts for Further Information

If you require any further information regarding this questionnaire or your participation in the study, you may contact Nasrin Ramezanali at [redacted] The interview will take approximately 10 to 15 minutes of your time. Those participants willing to participate will be asked the following questions:

1. How old are you and how long have you been studying English?
2. What strategies do you use to learn vocabulary?
3. What strategies do you use to remember vocabulary?
4. What is your opinion about vocabulary learning through definition-alone?
5. What is your opinion about vocabulary learning through definition and audio pronunciation?
6. What is your opinion about vocabulary learning through definition and video/animation?

7. Which vocabulary learning technique (definition alone, definition and audio pronunciation OR definition and video/animation) did you prefer/like, and why?
8. Did the definition alone mode help you remember the words in the final test? If no, why not? If yes, in what ways?
9. Did the definition and audio pronunciation mode help you remember the words in the final tests? If no, why not? If yes, in what ways?
10. Did the definition and video/animation mode help you remember the words in the final tests? If no, why not? If yes, in what ways?
11. How have the vocabulary learning modes practiced here changed the way you used to learn new words?
12. What did you like most about the vocabulary learning practice here?
13. What did you like least about the vocabulary learning practice here?
14. What other thoughts do you have to share with me about vocabulary learning by providing text, audio and video definitions?

Note: Additional follow-up questions will be asked, as appropriate, with each participant.

Appendix R: Tables of Assumptions

Table R.1 Tests of Normality for VLT Data

	Group	Kolmogorov-Smirnov[a]		
		Statistic	df	Sig.
VLT Pre-test	Gr. A (Control)	.30	24	.000
	Gr. B (TAV)	.22	39	.000
	Gr. C (AVT)	.22	36	.000
	Gr. D (VTA)	.21	33	.001

a. Lilliefors Significance Correction

Table R.2. Levene's Test of Equality of Error Variances[a] (PR test)

Dependent Variable: Immediate. Post-test

F	df1	df2	Sig.
5.624	3	103	.001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Pretest + Group

Table R.3 Levene's test of equality of error variances[a] (PR test)

Dependent variable: delayed post-test

F	df1	df2	Sig.
4.01	3	99	.010

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Pretest + Group

Table R.4 Levene's Test of Equality of Error Variances (PR test)

	F	df1	df2	Sig.
Pre1-im1	8.618	3	115	.000
Pre2-im2	3.709	3	111	.014
Pre3-im3	7.801	3	110	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Table R.5 Levene's Test of Equality of Error Variances (PR test)

	F	df1	df2	Sig.
im1-del1	1.819	3	101	.149
im2-del2	1.386	3	98	.252
im3-del3	5.964	3	98	.001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Table R.6 Levene's Test of Equality of Error Variancesa (PR test)

Dependent variable: sub delayed post-tests				
	F	df1	df2	Sig.
Sub-test 1	1.788	3	110	.154
Sub-test 2	7.050	3	109	.000
Sub-test 3	1.557	3	99	.205

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Table R.7 Levene's Test of Equality of Error Variances[a] (MC Test)

Dependent variable: immediate post-test			
F	df1	df2	Sig.
.035	3	103	.991

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Pretest + Group

Table R.8 Levene's test of equality of error variances[a] (MC Test)

Dependent Variable: Delayed post-test			
F	df1	df2	Sig.
4.564	3	109	.005

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Pretest + Group

Table R.9 Levene's Test of Equality of Error Variances[a] (MC test)

	F	df1	df2	Sig.
Pre1-im1	.219	3	115	.883
Pre2-im2	1.513	3	111	.215
Pre3-im3	3.607	3	110	.016

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Pre.Sub1 + Group

Table R.10 Levene's Test of Equality of Error Variances (MC Test)

	F	df1	df2	Sig.
Im1-del1	2.140	3	101	.100
im2-del2	1.828	3	99	.147
im3-del3	2.174	3	103	.096

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Table R.11 Levene's Test of Equality of Error Variancesa (MC test)

Dependent variable: sub delayed post-tests

	F	df1	df2	Sig.
Sub-test 1	1.788	3	110	.154
Sub-test 2	1.050	3	109	.205
Sub-test 3	7.557	3	99	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Appendix S: Ethical Approval for Pilot Study

Western Research

Research Ethics

Western University Health Science Research Ethics Board
NMREB Delegated Initial Approval Notice

Principal Investigator: Dr. Farahnaz Faez
Department & Institution: Education\Faculty of Education,Western University

NMREB File Number: [redacted]
Study Title: Investigating the Effectiveness of Simultaneous Multimedia Glossing on L2 Learners' Vocabulary Learning and Long-term Word Retention
Sponsor:

NMREB Initial Approval Date: December 17, 2014
NMREB Expiry Date: December 17, 2015

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Revised Letter of Information & Consent	Stage 2, ethics' letter of information and consent form (PDF format)	2014/12/09
Revised Letter of Information & Consent	Stage 1, ethics' letter of information and consent form (PDF format)	2014/12/09
Revised Western University Protocol	Revised Ethics' application form (PDF format)	2014/12/09

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the above named study, as of the NMREB Initial Approval Date noted above.

NMREB approval for this study remains valid until the NMREB Expiry Date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario.

Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number [redacted]

Ethics O[redacted] Chair or delegated board member

Ethics Officer to Contact for Further Information

[redacted]

This is an official document. Please retain the original in your files.

Western University, Research, Support Services Bldg., Rm. 5150
London, ON, Canada N6A 3K7 t. 519.661.3036 f. 519.850.2466 www.uwo.ca/research/services/ethics

Appendix T: Ethical Approval for Actual Study

Western Research

Research Ethics

Western University Non-Medical Research Ethics Board
NMREB Delegated Initial Approval Notice

Principal Investigator: Dr. Farahnaz Faez
Department & Institution: Education\Faculty of Education,Western University

NMREB File Number: [redacted]
Study Title: Investigating the Effectiveness of Simultaneous Multimedia Glossing on L2 Learners' Vocabulary Learning and Long-term Word Retention: A Mixed Methods Research
Sponsor:

NMREB Initial Approval Date: September 08, 2015
NMREB Expiry Date: September 08, 2016

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Revised Western University Protocol		2015/08/01
Revised Letter of Information & Consent		2015/08/01
Instruments	List of Words & Definitions	2015/08/01
Instruments	Revised demographic information form	2015/08/01
Instruments	Revised questionnaire	2015/08/01
Instruments	Revised semi-structured interview	2015/08/01
Instruments	Vocabulary Levels Test	2015/08/01
Instruments	Productive Recall Vocabulary test	2015/08/01
Instruments	Multiple choice Productive Recognition vocabulary Test	2015/08/01
Instruments	Reading Text 1	2015/08/01
Instruments	Reading Text 2	2015/08/01
Instruments	Reading Text 3	2015/08/01

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the above named study, as of the NMREB Initial Approval Date noted above.

NMREB approval for this study remains valid until the NMREB Expiry Date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario.

Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The NMREB [redacted]ealth & Human Services under the IRB registration number [redacted]

Ethics Offic[redacted] or delegated board member

[redacted]icer to Contact for Further Information

[redacted]

This is an official document. Please retain the original in your files.

Western University, Research, Support Services Bldg., Rm. 5150
London, ON, Canada N6G 1G9 t. 519.661.3036 f. 519.850.2466 www.uwo.ca/research/ethics

CURRICULUM VITAE

NASRIN RAMEZANALI

EDUCATION

Degree	Discipline	University	Year
Doctor of Philosophy (PhD).	Curriculum Studies & Studies in Applied Linguistics	Western University	2017
Master of Art	English Language Teaching (ELT)	University of Al-Zahra Iran	2004
Bachelor of Art	English Language Teachings (ELT)	University of Arak Iran	2000

SCHOLARSHIPS AND AWARD

Year	Scholarship/Award	Amount
May 2017	Faculty of Education Conference Travel Program	553.68 $
June 2016	Faculty of Education Conference Travel Program	658.25 $
September 2016	Faculty of Education Conference Travel Program	327.31 $
April 2016	Faculty of Education Conference Travel Program	553.29 $
May 2016	Student Research Grant	294.78 $
July 2015	Student Research Grant	1660.13 $
August 2014	Textbook Grant	201.44 $
April 2014	Society of Graduate Students (SOGS) Travel Award	105 $

RESEARCH EXPERIENCE

Date	Position	Institution
September 2014 - April 2015	Research Assistant PI: Dr. Farahnaz Faez Research Project : Corpus Linguistics Technical Term-extraction Terminology Extraction	Western University
September 2015- May 2017	Graduate Research Assistant PI: Dr. Farahnaz Faez Research Project : Vocabulary glossing & Multimedia learning	Western University

SELECTED CONFERENCE PRESENTATIONS (REFEREED)

Ramezanali, N., & Faez, F. (2017). Short and Long-term Vocabulary Retention through Multimedia Glossing: A Mixed Methods Research. Paper Accepted for presentation at EuroSLA 2017, Reading, UK.

Ramezanali, N., & Faez, F. (2017). Word Retention through Multimedia Glossing: A Mixed Methods Research. Paper presented at Canadian Association of Applied Linguistics (CAAL/ACLA), Toronto, ON, Canada.

Ramezanali, N., & Faez, F. (2016). Multimedia Glossing and Vocabulary Learning: A Mixed Methods Research Examining Word Retention. Paper Presented at Second Language Research Forum (SLRF), Columbia University, New York.

Ramezanali, **N.,** & Faez, F. (2016). Multimedia glossing, vocabulary acquisition and long-term word retention: A Mixed methods research. Paper presented at Association of American Applied Linguistics (AAAL), Orlando, Florida, USA.

Ramezanali, **N.** (2016). Effectiveness of multimedia glossing and long-term word retention: A Mixed methods research. Paper presented at Canadian Association of Applied Linguistics (CAAL/ACLA), Calgary, Alberta, Canada.

Ramezanali, **N.** (2015). Glossing and Vocabulary Learning, Canadian Association of Applied Linguistics (CAAL/ACLA), Round table presentation, Fairmont Royal York, Toronto, ON.

Ramezanali, **N**. (2015). The Impact of Multimedia Glossing on Vocabulary Retention, Robert Macmillan Graduate Research in Education Symposium, Round table presentation, Western University, London, ON.

Ramezanali, N. (2014). The Effect of Four Different Modes of Instruction on L2 Vocabulary Acquisition of EFL Learners. Paper presented at Canadian Association of Applied Linguistics (CAAL/ACLA), Brock University, St. Catharines, ON.