THE JOURNAL OF ASIA TEFL
Vol. 18, No. 1, Spring 2021, 1-22
http://dx.doi.org/10.18823/asiatefl.2021.18.1.1.1

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

Input Modality and Working Memory Capacity: Effects on the Acquisition of Receptive and Productive Vocabulary Knowledge

Ali Akbar Ansarin
University of Tabriz, Tabriz, Iran

Yaser Hadidi
University of Tabriz, Tabriz, Iran

Salva Kazemipour Khabbazi*
University of Tabriz, Tabriz, Iran

Research on relationships between textual and textual-plus-pictorial gloss types and working memory capacity has been limited at best. This study examined the degree to which these two gloss types and working memory capacity (WMC), assessed through a listening span task mediate vocabulary learning. 204 English as a Foreign Language (EFL) learners listened to a text with an unfamiliar topic, before and after which they completed a vocabulary test based on Paribakht and Wesche's (1997) Vocabulary Knowledge Scale to examine their receptive versus productive vocabulary acquisition. While listening to the text, they received the meanings of twenty selected words through two different gloss types. The results revealed that textual-plus-pictorial glosses led to significantly higher increases in both receptive and productive knowledge. As for the predictive power of working memory capacity, linear regression analysis indicated that WMC did not play a role in receptive vocabulary acquisition in either of the input modes, whereas productive gain scores for both gloss types were equally, albeit insignificantly, predicted by WMC. However, no interaction between WMC and the modality effect was observed. This study confirms the modality effect on intentional vocabulary learning and suggests that the role of WMC is not mediated by gloss types.

Keywords: working memory capacity, modality effect, gloss type, receptive vocabulary knowledge, productive vocabulary knowledge.

Introduction

Working memory capacity, as one of the key cognitive sources for individual differences (Skehan, 2012), is believed to be a contributing factor in second language acquisition (SLA) (Mackey & Sachs, 2012; Révész, 2012). Empirical studies have shown a positive correlation between WMC and performance in TOEFL, second language (L2) proficiency outcomes, bilingualism, grammar and vocabulary learning, reading and listening comprehension plus writing abilities (Denhovska, Serratrice, & Payne, 2016; Grundy & Timmer, 2016; van den Noort, Bosch, & Hugdahl, 2006). However, learning conditions, of which input mode is one manifestation, are also influential in mediating the interaction between WMC with different aspects of language learning. Cognitive Theory of Multimedia Learning (CTML) proposes the use of audiovisual presentations over visual-only presentations to reduce the load

on WMC, and thus increase acquisition (Inan et al., 2013; Mayer, 2014). As a result, several studies (Schüler, Scheiter, & van Genuchten, 2011) have investigated the use of different gloss types in multimodal instruction. So far, the main focus of research has mainly been on the effect of gloss types on reading and reading comprehension with only a few studies investigating this effect on listening. Moreover, only a small number of studies have examined the effect of gloss types on vocabulary learning in listening comprehension (Çakmak & Erçetin, 2017).

Previous research has indicated that despite the usefulness of multimodal instruction in some areas, not all gloss types are effective in second language (L2) vocabulary learning (Çakmak & Erçetin, 2017; Cottam, 2010). In exploring the factors affecting vocabulary learning, the role of WMC has figured prominently in theory and research. The focus, however, has been primarily on vocabulary learning in reading a text (e.g., Yang, Shintani, Li, & Zhang, 2017). Nonetheless, as a platform to process new information, WMC also influences listening to a great extent. It can, therefore, be assumed that WMC is a contributing factor in vocabulary learning through listening too. However, research is yet to address the gaining of vocabulary knowledge through glosses in listening and the role of WMC in this regard.

In this spirit, the aim of this study was twofold. The first objective was to verify and extend previous research on the role of multimodal instruction in L2 vocabulary learning. To be more precise, the focus of this study was to explore the effect of two gloss types on both receptive and productive vocabulary knowledge through listening. The second objective was to shed light on the predictive power of WMC in vocabulary learning through: a) different gloss types and, b) listening to an audio text.

Literature Review

Modality Effect and Vocabulary Learning

One of the questions addressed in this study was to explore the extent to which the modality effect influenced the vocabulary learning of L2 learners with different working memory capacities. Due to the presumed relationship between the modality effect and WMC (Schüler et al., 2011), one particular concern of the present study was to investigate this relationship further for proficient EFL learners' retention of L2 vocabulary through a multimodal expository presentation.

According to Cognitive Load Theory (CLT), WM is where the selection, organization, and integration of visual and auditory information occurs and the load imposed on WM resources while performing a task is referred to as Cognitive Load (CL) (Sweller, Ayres, & Kalyuga, 2011). The modality effect or multimodal instruction is one of the techniques employed to reduce the load on WM, thereby preserving more WM resources for the most optimal processing of incoming information (Sweller, 2010). It refers to the presentation of information by means of integrating different types of modes, namely, presenting verbal information either visually or auditorily, and pictorial presentation via static pictures, illustrations, graphs, charts, and diagrams or dynamic animations and videos (Mayer & Moreno, 2003). Compared to reading a text while viewing pictures, the modality effect emphasizes the presentation of materials in an audiovisual manner (Mayer, 2009).

Most studies have indicated the advantages of the modality effect, with some exceptions (Tabbers, Martens, & van Merriënboer, 2004). The intrinsic load, which is influenced by the inner complexity of the task, has been reported as a source of difference in these studies. In this respect, the modality effect increases when the intrinsic load of the task is high (Kozan, 2016). Furthermore, it has been stated that this effect is moderated by the pace of presentation. In self-paced instruction, the participants control the pace of the whole process of reading/listening together with access to the glosses. In system-paced instruction, on the other hand, the pace of the presentation, that is, both the reading/listening text and the glosses, is determined by the system over which the participants have no control.

Many studies have favored system-pacing over self-pacing instruction (Schmidt-Weigand, Kohnert, & Glowalla, 2010), while some found no modality effect under self-pacing (e.g., Wouters, Paas, & van

Merriënboer, 2009), or found a 'reverse modality effect' instead (better learning outcomes in visual presentation of the text rather than in auditory presentation) when the instruction was self-paced (Tabbers et al., 2004). These findings demonstrate the stronger effects of system-paced materials over self-paced ones.

The studies in this field, however, have focused on the effect of multimodal instruction on reading comprehension, with a comparatively small number of them investigating the relationship of multimodal instruction and vocabulary learning in a reading context (e.g., Ramezanali & Faez, 2019; Türk & Erçetin, 2012; for a review of previous research, see Yun, 2011). Equally, there have been fewer studies on the effect of multimodal instruction on vocabulary learning through listening (Cottam, 2010; Çakmak & Erçetin, 2017). The results found in these studies have been contradictory, with some showing a stronger effect for textual-plus-pictorial glosses over other types of glosses such as textual, video or animation, and multiple choice glosses, and some showing no difference in the effect of gloss types on vocabulary learning (e.g., Wang & Lee, 2020; Yoshii, 2013). One of the aims of the present study was to further explore this unresolved gloss type issue.

Working Memory Capacity and Vocabulary Learning

Another question addressed in this paper was whether WMC mediates the amount of L2 vocabulary learning through listening to an audio text while being presented with textual and textual-plus-pictorial gloss types. WMC is the system of storing chunks of information while performing other tasks such as comprehension, learning or reasoning (Baddeley, 2015) and while facing other processing or distractions (Conway et al., 2005). Previous research has implied that learners' acquisition and performance in L2 are accounted for by their working memory capacities (Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Speciale, N. Ellis, & Bywater, 2004). More specifically, some studies have found an above-average correlation between learners' WMC and their vocabulary development (French, 2006).

The studies in this domain have investigated either phonological short-term memory (PSTM) or executive working memory. According to the insight these studies offer, the essential feature distinguishing executive working memory from PSTM is that the former involves both storage and executive (process) functions of memory (Daneman and Carpenter 1980; Conway et al. 2005), while the latter does not involve either function. PSTM is measured by digit or word span tasks; executive working memory, however, is measured using sentence, operation, or counting span tasks that involve both storing and processing information (Li, 2017; Wen, 2016). Most studies have argued that PTSM is the predictor of L2 vocabulary learning in initial stages of learning and in decontextualized settings (Engel & Gathercole, 2012; Martin & Ellis, 2012), while for more proficient learners, executive working memory has been shown to play the chief role in vocabulary learning (Yang et al., 2017).

Since an important role has been attributed to WMC in the processing of L2 verbal input and verbal intelligence (Nowbakht & Fazilatfar, 2019), most of the studies in this field have primarily been concerned with WMC as a facilitator of comprehension of the content of L2 texts (Joh, 2018; Joh & Plakans, 2017; Kozan, Erçetin, & Richardson, 2015). Relatively few studies have explored the effect of WMC on learning vocabulary or grammatical forms besides comprehension in contextualized settings (Leeser, 2007; Yang et al., 2017); the findings are indicative of a positive relationship between WMC and the learning of vocabulary and syntax. Notably, the common point in these studies is that vocabulary or syntax learning was examined through reading an L2 text, while intentional vocabulary or grammatical learning through listening seems to have been neglected. That is to say, to the best of our knowledge, all the studies conducted in this area have focused on the relation between WMC and incidental vocabulary and grammar acquisition (Denhovska et al., 2016; Leeser, 2007; Yang et al., 2017). In this spirit, the present study set out to fill this gap in the literature and to examine the role WMC plays in intentional vocabulary development, which is an important area yet to be explored (Guzmán Muñoz, 2018). The fact that previous research has not related intentional vocabulary learning to WMC as much as it did with incidental vocabulary learning is rather counter-intuitive. This is because incidental vocabulary learning

is the learning of one feature in the incoming stimulus signal while concentrating on another feature simultaneously, whereas intentional learning is defined as having the intention to learn the material and to commit it to one's memory (Schmitt, 2010; Webb, 2020). Along the same lines, the impetus feeding into the current research is the fact that intentional vocabulary acquisition and its manifestation in listening comprehension involve an actual intentional focus on word–meanings, while in incidental vocabulary learning, intake is based on contextual clues (Lindstromberg, 2020). The reason why research has looked more extensively into incidental vocabulary learning was the repeated findings that learners guess the meanings of the new words through the context more easily, as well as the fact that incidental vocabulary learning has been argued to motivate learners for extensive reading (Choi, 2016; Chun, Choi, & Kim, 2012; Waring & Takaki, 2003). However, the interaction of gloss types, listening comprehension, and the incidental/intentional dichotomy is yet to be more fully explored.

As such, in addition to the unresolved gloss type issue, the role of WMC in vocabulary learning through a listening text as opposed to reading has not been thoroughly examined. To our knowledge, no research has addressed the amount of vocabulary learning in a multimodal environment (that is, through different gloss types) in combination with the powers of WMC to predict both receptive and productive vocabulary learning in such an environment. With the aim of filling the abovementioned gap and shedding more light on this area, this study investigated the relationship between WMC (i.e. executive working memory) and intentional vocabulary learning through a multimodal instruction in which an audio text was accompanied by two gloss types for the intended words.

Method

Research Questions

In order to address identified gaps, this study attempted to answer the following questions:

1. Do different glossing modes (textual vs. textual plus pictorial glosses) vary in their efficacy in receptive and productive word knowledge of EFL learners?
2. Does working memory capacity predict EFL learners' gain scores of receptive and productive word knowledge for words presented through different glossing modes (textual vs. textual plus pictorial glosses)?

Participants

The participants were 204 students (146 females and 58 males; M = 20.47 years) of 6 intact classes who were majoring in English language and literature at three public universities in East-Azerbaijan, Iran (two classes in each university). These were the three highest-ranking universities in the region where we considered junior and senior undergraduate students for participation in the study. They were all enrolled at the Department of English where English is the medium of instruction. Based on their high scores (beyond 75 percent) in the special English language test of the National University Entrance Examination, they can be considered as proficient English learners (B2 level or independent users according to CEFR). This special English language test consists of vocabulary, grammar, reading comprehension, language function, and cloze modules at an advanced level. It is worth mentioning that, initially, there were 245 students, some of whom dropped out during the study (n = 18), and some of whom were excluded from the experiment as their z-scores were beyond the normal range (n = 23). The participants were informed of the immediate vocabulary post-test.

Design

As an experimental study, a within-subjects design was adopted. The study lasted about 3 weeks for each of the 6 classes. There was a vocabulary pre-test in week 1, a working memory test in week 2 so that the equally-distanced phases of the experimental method from the baseline and before the treatment could be observed, and finally a treatment and an immediate vocabulary post-test in week 3. As the dependent variable of the study, the participants' vocabulary knowledge was measured in terms of receptive and productive knowledge. Unlike most studies which consider form recall entailing the provision of the L2 equivalent of the word as the productive knowledge of words (Çakmak & Erçetin, 2017; Eckerth & Tavakoli, 2012; Webb & Kagimoto, 2009), productive vocabulary knowledge, in the present study, was indicated by the use of the target words appropriately in a sentence.

The vocabulary items were presented through a listening text in two different gloss modes of word meaning or word meaning + picture (textual vs. textual plus pictorial glosses). Gloss types and WMC were independent variables of the study. It is worth mentioning that all participants were presented with both of the gloss types, hence a within-subject design for this independent variable. In all stages of the research, the participants completed the tests individually with no discussions about the target words with other participants and no access to dictionaries or any digital devices.

Instructional Materials

The treatment used was the listening text on Page 172 [mp3 039-040] from the Third Edition of Longman Preparation Course for the TOEFL iBT Test, (Phillips, 2015). The listening passage is a lecture in a zoology class about opossum, particularly its defensive mechanisms. In line with the sampling and selection practice of major research in this area (e.g., Çakmak & Erçetin, 2017; File, 2010), twenty-five vocabulary items were selected from this passage which were the maximum possible number of most demanding items in this text for these students in lieu of their low-frequency and possibly least familiar nature. Following a pilot test with similarly proficient participants, five vocabulary items were discarded, since the participants indicated familiarity with these items. The remaining twenty items were divided into two glossed groups of text only (i.e., the Persian translation of the word), and text + picture (i.e., the Persian translation plus a picture representing the meaning of the word). The ten items chosen for text only group were bluff (v.), claw (n.), crawl (v.), foul (adj.), hiss (v.), intimidation (n.), project (onto) (v.), reiterate (v.), snarl (v.), stiffen up (v.). The other ten items for text + picture group were cling (v.), emit (v.), marsupial (n.), opossum (n.), outrun (v.), play possum (phr.), pouch (n.), racket (n.), stench(n.), venom (n.). The rationale for this categorization was the words' potential to lend themselves lucidly to depiction through pictures. More abstract words, not as easily and unambiguously amenable to pictorial depiction, were selected for the text-only group. The appendix section contains all the glosses used in the treatment session. All the vocabulary items were checked for their frequency, none being on the 2000-frequency-word lists; that is, the Academic Word List (AWL) and University Word List (UWL), suggesting, subsequently, that they could be considered low-frequency words.

The listening text was accompanied by visual slides that involved either the Persian translation of the words alone or the translation plus a picture. They appeared the instant the Zoology professor mentioned them. As a system-paced condition was put in place in this study, each of the slides remained on the screen for 4 seconds. The aim was to measure the amount of vocabulary learned while listening to a text accompanied by visual aids for the target words.

Testing Instruments

Immediately after the listening text, the participants answered five multiple-choice listening comprehension questions, by way of analyzing the amount of attention given to the listening text besides the vocabulary slides. As Brunfaut and Révész (2014) maintain, comprehension questions that assess

listeners' local comprehension relate to WMC more strongly than those tests measuring global comprehension. As a result, the comprehension questions in this study assessed local comprehension by asking for details in the listening text.

Vocabulary test

As Schmitt (2010) and Nation (2001) assert, learners tend to move along a continuum in their knowledge about vocabulary items, some aspects being learned before the others. Moreover, in order to use a language efficiently, plus a large vocabulary size, knowledge of lexical items is also required (Schmitt, 2010). It means that a form-meaning link, probably sufficient for recognition, will most likely not be enough for productive uses. Therefore, a test was needed to cover different levels of vocabulary knowledge; that is, to recognize the meanings of words as well as to use them in a sentence. In line with this purpose, Paribakht and Weshche's (1997) Vocabulary Knowledge Scale (VKS) was employed in order to measure the participants' receptive and productive knowledge of the target words before and after the treatment. This test, already used in many studies on vocabulary development (File & Adams, 2010; Kim, 2011; Yang et al., 2017), has a 5-point scale in which self-perceived knowledge and actual performance are scored, tracking both the receptive and productive development of specific word knowledge in learners (Read, 2000).

In these five points on this scale, the first and second categories (i.e., 1: "I do not remember having seen this word before" and 2: "I have seen this word before, but I don't know what it means") require the participants only put a tick mark in the box without any answers needed. For the third, fourth and fifth categories (i.e., 3: "I have seen this word before, and I think it means", 4: "I know this word and it means", 5: "I can use this word in a sentence"), the participants have to demonstrate their knowledge by providing a synonym or the translation of the words (categories 3 and 4), and a sentence with the target words (category 5).

Both the pre- and post-tests were given in English; the instructions, however, were provided orally by the researcher in Persian to ensure the participants' complete understanding of the procedures. The tests consisted of the 20 target words plus 4 distractors. The order in which the words appeared in the pre-test and post-test was randomized to prevent a carry-over effect from the previous sessions. Prior to the treatment, the participants were informed about the imminent post-test.

Working memory test

Most relevant SLA research has made use of operation span tasks or reading span tasks (Li & Fu, 2018; Miller, Fox, Moser, & Godfroid, 2018; Yang et al., 2017). These two tests, however, seemed inappropriate for our purpose for two reasons. Firstly, operation span tests tap into aspects of WMC that are unrelated to language. Since this study aimed at finding the relationship between WMC and L2 vocabulary learning, using this test would have led us to misleading results. Secondly, as listening and reading are two different skills in language learning, the reading span test might have led to a misinterpretation of the results. Thus, in order to address the research question to do with probing language-related aspects of WMC via listening tasks, a listening span test was considered to be an appropriate measure in this study.

The selected working memory test was the spoken version of the reading span test (Daneman & Carpenter, 1980) which requires the individuals to listen to some sentences, judge their semantic plausibility (i.e., process) and memorize the last word of the sentences (i.e., storage) (Mackey, Adams, Stafford, & Winke, 2010). The audio test (Daneman & Carpenter, 1980) consisted of 60 English sentences (L2) presented orally. The sentences were randomly divided into 15 sets of 2 to 6 sentences. The sets were randomly ordered so that a set with 2 sentences, for instance, was not necessarily followed by a set with 3 sentences but was followed by a set with 5, 6 or 4 sentences. In addition, none of the final words were semantically related to each other, in an attempt to prevent the participants from forming a

meaningful connection among the words which would have contributed to their memorization. As the test was administered in English, some sentences with difficult final words were replaced with other easier sentences of the test to exclude listening ability as a confounding factor in the study.

After instructions given in Persian (L1) and three practice sets to get familiar with the test, the participants listened to the sentences and made semantic judgments while trying to memorize the final word of each sentence. There was a time interval of 2 seconds between the sentences to give the participants time to make the judgments and put a check mark on their answer sheets. As the sentences were numbered both in the audio file and the answer sheet, immediately after the completion of the judgment task for the sentences, the participants could start recalling and writing the final words of all sentences of the set. They had 5 seconds for each sentence in the set, such that if there were 4 sentences in the set, they were given 20 seconds to recall and write the final words. The participants were told that writing the words in their order of appearance was not necessary, so they could write the final words in the order in which they recalled them without this procedure having any effect on their scores (Chen, 2013). At the end of the writing period, there was a beep after which the new set would start.

Scoring and Data Analysis

Performance on vocabulary test

The scoring scheme for EFL learners' knowledge of the target words was based, in part, on Paribakht and Wesche's (1997) and File and Adam's (2010) scoring systems. Given the purpose of our study which was to examine the differences caused by the treatment on the participants' receptive and productive knowledge, the scoring was divided into two parts. One score was given to receptive knowledge (category 1-4 in the test) and one to the productive one (category 5 in the test). For receptive knowledge of the words, as the first category showed no such knowledge, there were no points. The second category received one point although the participants did not provide the meaning. Two and three points were given to the third and fourth categories, respectively.

Since any changes, even the minor ones, between the pre-test and post-test, were meant to be measured, different points were given to category 3 and 4 even when the given answer was correct. Accordingly, when a participant's answer was correct but was written in the third category ("I think it means"), 2 points were given as opposed to the 3 points for the fourth category ("I know this word and it means"). Therefore, the participants' degree of certainty about the meaning of the items changed the scores they received. In addition, if an incorrect answer was provided in the fourth category, only two points were awarded. Maximum score for the receptive part of the test was 60 (30 for each gloss type).

As for the productive evaluation of the participants' knowledge, if no sentences were provided at all, zero point was given. If the written sentence was semantically appropriate, four points were given while the semantically and grammatically correct one received a score of five. As the participants were relatively proficient EFL learners, simple sentences, which were indicative of superficial knowledge of the words, did not receive any scores. Maximum obtainable score for this part of the test was 100 (50 scores for each gloss type).

It should be mentioned that two researchers scored VKS tests (the pre- and post-tests) for all of the 204 participants. Inter-rater reliability was high (95%) and all the disagreements were deliberated on until agreement was reached. Hence, the scores for receptive and productive knowledge were added up in two distinct parts and the gain scores for both were calculated to be further analyzed.

Performance on listening span test

Scores on the working memory test were calculated based on the partial-credit load scoring procedure (Conway et al., 2005). According to this procedure, each correctly recalled item was awarded one point

regardless of its position within a set and regardless of the number of items in the set. No point was given to the item recalled incorrectly or not recalled at all.

As for scoring the processing part of the test (i.e., semantic judgments), Conway et al. (2005) argued that there is no need to calculate the correctness of judgments, since the accuracy of judgments is usually close to the ceiling due to the emphasis placed on attending to the task. This reason is coupled with the positive correlation observed in most studies between the performance on the processing task and recalling task, leaving us with no evidence of processing versus storage trade-offs (Kane et al., 2004). Thus, points can be given to the recalled items even if the processing task was not done accurately. As a result, under the condition of having 85% accuracy on the judgment task, the data for the subject's performance on the storage part of the task were kept in the analyses (the judgment score of all the participants was beyond 85% accuracy).

Data analysis. The participants' scores on two VKS tests, and their working memory scores were calculated by SPSS version 23. In order to compare the effect of gloss types on receptive and productive word knowledge, a paired-samples t-test was conducted. A regression analysis was conducted to examine the effect of WMC on the participants' vocabulary development.

Results

Gloss Types and Receptive and Productive Vocabulary Knowledge

The first research question in this study concerned the effect of two different gloss types on receptive and productive vocabulary learning. As a pre-test and a post-test were given to the participants, gain scores for the difference of post- and pre-tests were calculated for each gloss type. That gave us four sets of scores for each participant, that is, receptive scores for gloss type 1 and 2, plus productive scores for gloss type 1 and 2. Table 1 shows the descriptive and inferential results for the vocabulary and WM tests.

A paired-samples t-test was conducted to compare the learners' receptive and productive vocabulary scores in textual and textual-plus-pictorial gloss types. The results are mapped out in Table 2. They indicate significant differences in receptive and productive scores between the two gloss types. There was a significant difference in the receptive scores for textual (M = 11.92, SD = 5.24) and textual-plus-pictorial (M =15.50, SD =5.09) glosses; $t(203)$ = -10.98, p = 0.0001 with textual-plus-pictorial glosses leading to more learning than textual glosses.

The null hypothesis stating that the gloss types do not have any effects on the receptive scores of the participants was rejected. Further, Cohen's effect size value (d = 0.69) suggests a moderate to high practical significance. Comparably, in the scores for productive vocabulary knowledge between textual (M = 22.08, SD = 11.76) and textual-plus-pictorial (M = 28.29, SD = 11.44) glosses, a significant difference ($t(203)$ = -8.65, p = 0.0001) was observed.

TABLE 1
Descriptive Statistics for Vocabulary and WM Tests

	Receptive Knowledge		Productive Knowledge		WMC	
	M	*SD*	*M*	*SD*	*M*	*SD*
Textual Glosses	11.92	5.246	22.08	11.767	31.14	7.922
Textual+Pictorial Glosses	15.50	5.091	28.29	11.443		

N = 204

TABLE 2
T-test Results of the Differences between Gloss Types

	n	*Mean*	*SD*	*t*	*df*	*Sig. (2-tailed)*	*d*
Pair 1 Receptive, gloss type 1, 2	204	-3.588	4.665	-10.987	203	.000	.692
Pair 2 Productive, gloss type 1, 2	204	-6.216	10.259	-8.653	203	.000	.528

$p < .05$

Consequently, the null hypothesis that gloss types do not differ in their efficacy of productive vocabulary knowledge with a moderate effect size (Cohen's $d = 0.52$) can be rejected. Similar to receptive vocabulary knowledge, textual-plus-pictorial glosses led to more learning rates than textual glosses. These results are suggestive that gloss types (textual and textual-plus-pictorial glosses in the present study) do not have the same effect on vocabulary learning through a listening text.

Working Memory Capacity, Gloss Type, and Receptive versus Productive Vocabulary Knowledge

In order to predict the learners' receptive and productive vocabulary gain scores based on their WMC scores, simple linear regressions were calculated. Similar to the case in our first question, four sets of scores were formed for the second research question, based on receptive and productive gain scores of the pre- and post-tests along with two gloss types (i.e., receptive scores of gloss type 1 and 2, productive scores of gloss type 1 and 2).

As Table 3 indicates, for the receptive gain scores for the words presented through textual glosses, no correlation was confirmed between the vocabulary scores and WMC scores, as the Pearson's correlation coefficient was as follows: $r = .014$, $n = 204$, $p = 0.422 > \alpha = 0.05$. Therefore, we accepted the null hypothesis that the receptive vocabulary gain scores of textual gloss type and WMC are not correlated. As a result, linear regression equation cannot be stated for the prediction of receptive vocabulary knowledge of textual glosses based on WMC scores. Consistently, for the receptive gain scores of gloss type two (i.e., textual-plus-pictorial gloss type), no correlation could be found between vocabulary scores and WMC scores ($r = .095$, $n = 204$, $p = 0.088$). Consequently, a linear regression equation can similarly not be stated for the prediction of receptive vocabulary knowledge of textual-plus-pictorial glosses based on the participants' WMC scores. Overall, the results indicate that WMC could not predict participants' receptive vocabulary knowledge in either gloss type.

For the productive gain scores of words presented through two different gloss types, however, the results were contradictory. We reject the null hypothesis that WMC scores could not predict productive vocabulary learning through textual gloss type, for the Pearson correlation coefficient was $r = 0.263$, $n = 204$, $p = 0.0001 < \alpha = 0.05$. Correspondingly, the linear regression equation could be stated for the productive vocabulary learning in textual gloss type (i.e., the dependent variable) and WMC scores (i.e., the independent variable).

TABLE 3
Regression Results for the Predictive Power of WMC on Receptive and Productive Vocabulary Knowledge

	r	*p-value*	R^2	*F*	β	α	*Sig.*
Receptive, text	.014	.422	.0001	.038	.009	11.632	.845
Receptive, text+pic	.095	.088	.009	1.84	.061	13.600	.176
Productive, text	.263	.0001	.069	14.99	.391	9.917	.0001
Productive, text+pic	.361	.0001	.130	30.190	.521	12.073	.0001

$p < .05$

A significant regression equation was found ($F(1, 202) = 14.99$, $p < 0.001$), with an R^2 of .069. The participants' predicted receptive gain scores in textual gloss type were equal to 9.917 + 0.391 × (WMC scores) such that their vocabulary scores increased 0.391 for each point in the WMC test. Nevertheless, as $R^2 = 0.069$ which is much less than .7, the model proves to be weak in predicting the receptive vocabulary scores of textual glosses based on the WMC scores.

Equally, the productive vocabulary gain scores of textual-plus-pictorial gloss type could be predicted based on their WMC scores. Considering the fact that Pearson's correlation coefficient was: $r = 0.361$, $n = 204$, $p = 0.0001 < \alpha = 0.005$, the null hypothesis stating that the two variables have no relation was rejected. Accordingly, the linear regression equation for WMC as our independent variable, and productive gain scores of the words presented through textual-plus-pictorial gloss type as our dependent variable were: $F(1, 202) = 30.190$, $p < 0.001$), with an R^2 of 0.130. Hence, the predicted productive vocabulary gain score of the participants in two gloss type is equal to 12.073 + 0.521 × (WMC score), so the participants' productive vocabulary scores in this gloss type increased 0.521 for each point in the WMC test. Thus, the WMC was a predictor of productive vocabulary scores in gloss type two. As $R^2 = 0.130$, however, WMC was a weak predictor of the productive scores.

Working Memory Capacity and Listening Comprehension Questions

After listening to the text, all participants answered five comprehension questions to make sure that, in combination with the presented words, they had also focused on the content of the listening text. As the relationship between vocabulary learning and WMC was examined, it was reasonable to investigate listening comprehension with WMC as well. Hence, a simple linear regression was conducted to analyze the extent to which WMC mediated learners' listening comprehension. Table 4 shows that $r = .204$, $n = 204$, $p = .003 < \alpha = 0.05$; the null hypothesis stating that WMC and comprehension questions have no relationships is rejected. As Table 4 indicates, the linear regression equation for WMC as the independent variable and comprehension question scores as the dependent variable were: $F(1,202) = 8.735$, $p < 0.001$, with an R^2 of 0.041. The participants' predicted comprehension question scores were equal to 2.044 + 0.030 × (WMC scores), meaning that the comprehension questions score increased .030 for each point in the WMC test. It is apparent that, with an R^2 of 0.041, WMC is a weak predictor of comprehension questions.

To sum up, the results revealed the significant modality effect on both receptive and productive vocabulary acquisition of proficient language learners. They also point to the predictive power of WMC for productive vocabulary acquisition.

TABLE 4
Regression Results for the Predictive Power of WMC on Listening Comprehension Questions

	M	*SD*	*r*	*p-value*	R^2	*F*	*β*	*α*	*Sig.*
Comprehension Questions	2.97	1.15	.204	.003	.041	8.735	.030	2.044	.003

$p < .05$

Discussion and Conclusion

Significant findings were obtained in relation to the differences in vocabulary acquisition in the two gloss types. In both receptive and productive vocabulary acquisition, textual-plus-pictorial glosses led to more acquisition rate suggesting that the participants benefited significantly more from pictures accompanying the meanings of the words. These findings lend support to the modality effect (Mayer & Moreno, 2003; Mayer, 2009) and are in line with other studies finding the modality effect to be influential in acquisition rate (e.g., Kozan et al., 2015; Ramezanali & Faez, 2019; Türk & Erçetin, 2012). Our results,

however, run counter to the results in a number of other studies (Cottom, 2010; Çakmak & Erçetin, 2017; Tabbers et al., 2004).

A plausible explanation for this discrepancy is offered by differences in the pace of presentation in different studies. In these studies, self-paced systems were employed so that participants were able to scroll backward and forward when reading a text or listening to it, which turned out to have a reverse modality effect on learning (Wouters et al., 2009; Tabbers et al., 2004). Self-paced presentations provide the opportunity to go back to what went unnoticed, thus simultaneous processing does not actually occur in such presentations. The participants can fill their processing gaps by repeating the presentation, both the text and the glosses in some cases, thereby focusing on the tasks separately rather than simultaneously. In self-paced instructions, as a result, the differences in gloss types disappear since participants have the time and opportunity to learn the materials at their own pace. In the current study, a system-paced presentation was employed since the chief purpose was to explore the simultaneous processing of information (i.e., the listening text and target words); it is only in a system-paced presentation that participants' WMC determines the amount of processed information.

Another factor contributing to this inconsistency is the proficiency level of participants. In a meta-analysis, Abraham (2008) concluded that beginner learners seem to benefit from glosses less than advanced learners. This seems to be a question of the threshold level of vocabulary knowledge which is required before exposure to new words can be beneficial for learners (Schmitt, 2010). The participants in both Cottom (2010) and Çakmak & Erçetin (2017) were low proficient learners, which justifies the inefficiency of textual or textual-plus-pictorial gloss types for comprehension and vocabulary acquisition in these studies, whereas the participants of our study were high proficient learners who did benefit from textual-plus-pictorial glosses.

The results also showed that WMC was not a predictor of receptive vocabulary acquisition in a multimodal environment. In neither of the gloss types, a relationship between WMC and receptive gain scores of the participants was found. This suggests that all the participants could equally benefit from both textual and textual-plus-pictorial glosses despite the differences in their WM capacities. We speculate that acquisition of receptive vocabulary knowledge through a listening text with textual and textual-plus-pictorial glosses was not complicated enough to engage WMC of the participants (Eitel, Kühl, Scheiter, & Gerjets, 2014). It can be argued that the participants in this study were high-proficient learners, leading to a degree of automaticity in lexical processing having been formed as a result (Schmitt, 2010; Siyanova & Schmitt, 2008). This automaticity allows learners to process lexical items rapidly so that differences in WMC are masked.

On the other hand, according to the results, there exists a relationship between WMC and productive vocabulary knowledge. In both gloss types, WMC scores proved to be a predictor of the participants' productive gain scores. These results support, in part, previous research suggesting a relationship between WMC and vocabulary learning (Yang et al., 2017). According to Robinson (2007), with increases in task load, the WMC effect rises, a condition supported in this study. It was observed that as acquiring receptive knowledge of the words did not impose extra load on WMC, its effect diminished and participants with either low or high WM capacities could provide the meanings of the words in an equal fashion. For acquiring productive word knowledge, however, the participants' attention was divided into both the meaning of the words and their use in the sentences of the listening text. Consequently, only participants with higher WM capacities were able to simultaneously learn the meanings of the words while grasping how to use them in a sentence. As the WM assessed in this study was the executive working memory, the results complement previous research in the sense that executive working memory is an influential factor in productive vocabulary development of relatively proficient EFL learners.

Another finding in our study is the lack of any relationships between gloss types and WMC. In the regression analysis for productive word knowledge, WMC could predict the gain scores of the participants in both textual and textual-plus-pictorial glosses. The predictive power of WMC for textual-plus-pictorial glosses was just slightly more than that of textual glosses. This is indicative of the fact that WMC did not mediate the gloss type effect on vocabulary learning of proficient learners in this study.

This finding deviates from previous studies that have indicated the interaction of WMC and the modality effect (Brunyé, Taylor, Rapp, & Spiro, 2006; Gyselinck, Jamet, & Dubois, 2008; Kozan et al., 2015). One of the reasons for the disparity between the results of this study and those of previous research is that while their focus was on comprehension, ours was on vocabulary knowledge. Mixed findings on the effect of WMC on the modality effect and vocabulary learning resonate with the complexities of the way WMC may affect L2 vocabulary knowledge as well as the fact that we need more research-based insight into multimodal learning principles, which calls for more research before their full application in L2 learning environments is further known (Plass & Jones, 2005). The current study has sought to understand how WMC and the modality effect (i.e., two gloss types) influence the acquisition of receptive and productive word knowledge.

Overall, the modality effect turned out to be an effective factor in vocabulary acquisition for high proficient English learners. On the other hand, the results should be approached with caution given the weak predictive power of WMC for even productive vocabulary knowledge. A logical conclusion from these findings would be that the WMC effect is not absolute and several factors influence this effect, such as learners' characteristics, their proficiency, the WMC test employed, the vocabulary test, the amount of previous knowledge, and so forth.

Two limitations need to be addressed. One concerns the nature of WM test. There are some observations to the effect that WMC is language independent (Kane et al., 2004; Trofimovich, Ammar, & Gatbonton, 2007). However, as listening to a text in a foreign language is more challenging than reading a text, it seemed logical to administer the WM test in English so that the conditions in the test reflected those in the treatment session. Nevertheless, it is suggested that further research conduct the WM test in both first and foreign language or that both operation span tasks and listening/reading span tasks be used to ensure accommodation of the differences in their results.

Another point that should be mentioned is that although the pre-test ruled out the effect of previous knowledge of the words on the total score, this effect on the whole performance was not assessed. The participants with previous knowledge could have focused on other aspects of the words, thus outperforming those with no previous knowledge, without their WMC being an effective factor in this regard. This issue, ignored in almost all studies, might be one of the reasons why WMC was a weak predictor of vocabulary learning in this study. Still further research can undertake to probe this even further and at continued depth.

To conclude, previous research on the relationship between WMC and vocabulary learning has focused on decontextualized vocabulary learning, incidental vocabulary learning through reading, or merely receptive vocabulary knowledge. This study constitutes a significant complement to previous research in that it both investigated the effects of WMC and gloss types on intentional vocabulary learning in a multimodal environment, and it managed to differentiate their effects on the forming of receptive and productive vocabulary knowledge. This study carries direct pedagogical implications for EFL teachers and material developers concerned with using materials which are of reassuringly more promise and benefit to their students. Providing visual aids enhances vocabulary development even at a high-intermediate language proficiency. Of equally immediate importance is the consideration of individual differences, such as WMC, in choosing the materials and methods which would fit the learners' capacities and capabilities.

Acknowledgements

This paper is based on my doctoral dissertation (Salva Kazemipour Khabbazi, 2021)

The Authors

Ali Akbar Ansarin (PhD) is a professor of Applied Linguistics at the University of Tabriz, Tabirz, Iran. His areas of interest are Second Language Acquisition, Research Methods, and Psycholinguistics.

Department of English Language and Literature
University of Tabriz
29 Bahman Blvd.,Tabriz, Iran
Email: Ansalak@gmail.com

Yaser Hadidi (PhD) is an assistant professor of Applied Linguistics at the University of Tabriz, Tabriz, Iran. His research interests include Cross-linguistic probing of Conceptual Metaphors, Discourse Analysis, Literary Linguistics of Prose Fiction and Literature in Language Education. Two of his recent publications include 'Revisiting Common Source and Target Domains in Conceptual Metaphors in a Sample of English Fiction: Implications for Literacy Practices and Advanced EFL Pedagogy' and 'Exploring the Metaphors of Loyalty, Courage and Friendship in Harry Potter Novels and their Turkish Translations'.

Department of English Language and Literature
University of Tabriz
29 Bahman Blvd.,Tabriz, Iran
Email: hadidiy@tabrizu.ac.ir

Salva Kazemipour Khabbazi (corresponding author) completed her PhD in Applied Linguistics at the University of Tabriz in February 2021. She is currently a lecturer at Azarbaijan Shahid Madani University and the University of Tabriz. Her research interests include Psycholinguistics and Discourse Analysis.

Department of English Language and Literature
University of Tabriz
29 Bahman Blvd.,Tabriz, Iran
Email: Kazemipour_s@tabrizu.ac.ir

References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199–226. http://dx.doi.org/10.1080/09588220802090246.

Baddeley, A. D. (2015). Working memory in second language learning. In Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Bristol: Multilingual Matters.

Brunfaut, T., & Révész, A. (2014). The role of task and listener characteristics in second language listening. *TESOL Quarterly, 49*(1), 141–168. http://doi:10.1002/tesq.168.

Brunyé, T. T., Taylor, H. A., Rapp, D. N., & Spiro, A. B. (2006). Learning procedures: The role of working memory in multimedia learning experiences. *Applied Cognitive Psychology, 20*, 917-940. http://doi.org/10.1002/acp.1236.

Çakmak, F., & Erçetin, G. (2017). Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL, 30*(1), 24–47. DOI: http://doi.org/10.1017/S0958344017000155.

Chen, X. (2013). *Chinese EFL learners' noticing of recasts: Its relation to target structures, uptake, and working memory capacity* (Unpublished doctoral dissertation). Michigan State University, Michigan.

Choi, S. (2016). Effect of L1 and L2 glosses on incidental vocabulary acquisition and lexical representations. *Learning and Individual Differences, 45*, 137-143. doi:10.1016/j.lindif. 2015. 11.018

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. http://dx.doi.org/10.1016/j.neulet.2012.05.069

Conway, A., Kane, M., Bunting, M., Hambrick, D. Z., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769–786. https://dx.doi.org/10.3758/BF03196772

Cottam, M. E. (2010). *The effects of visual and textual annotations on Spanish listening comprehension, vocabulary acquisition and cognitive load* (Unpublished doctoral dissertation). Arizona, USA: Arizona State University. Retrieved from http://www.learntechlib.org/p/118131.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466. http://dx.doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Carpenter, P. A. (1980). Listening span task [Audio file]. Retrieved from https://www.millisecond.com/download/library/listeningspan.

Denhovska, N., Serratrice, L., & Payne, J. (2016). Acquisition of second language grammar under incidental learning conditions: The role of frequency and working memory. *Language Learning, 66*(1), 159–190. http://doi:10.1111/lang.12142

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*, 227–252. doi: 10.1177/1362168811431377

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: does harder-to-read mean better-to-understand? *Applied Cognitive Psychology, 28*, 488–501. http://doi.org/10.1002/acp.3004

Engel, M., & Gathercole, E. (2012). Executive and phonological processes in second language acquisition. *Journal of Educational Psychology, 104*, 974–986. http://doi:10.1037/a0028390

File, A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly, 44*, 222–249. http://doi:10.5054/tq.2010.219943

French, L. M. (2006). *Phonological working memory and L2 acquisition: A developmental study of children learning French in Qu´ebec*. Lewiston, NY: Edwin Mellen Press.

Grundy, J. G., & Timmer, K. (2016). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research, 33*(3), 325–340. http://doi:10.1177/0267658316678286

Guzmán Muñoz, F. J. (2018). The influence of personality and working memory capacity on implicit learning. *Quarterly Journal of Experimental Psychology, 71*(12), 2603-2614. https://doi.org/10.1177/1747021817749582

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology, 22*, 353–374. http://doi.org/10.1002/acp.1411

Inan, F. A., Crooks, S. M., Cheon, J., Ari, F., Flores, R., Kurucay, M., & Paniukov, D. (2013). The reverse modality effect: Examining student learning from interactive computer-based instruction. *British Journal of Educational Technology, 46*(1), 123–130. http://doi:10.1111/bjet.12129

Joh, J. (2018). Interplay of working memory, strategy use, and task difficulty in L2 reading comprehension. *The Journal of Asia TEFL, 15*(4), 900-914. http://dx.doi.org/10.18823/asiatefl.2018.15.4.1.900

Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System, 70*, 107–120. http://dx.doi.org/10.1016/j.system.2017.07.007

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial

memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217. http://doi:10.1037/0096-3445.133.2.189

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 61*, 100–140. http://doi.org/10.1111/j.1467-9922.2011.00644.x

Kozan, K. (2016). The incremental predictive validity of teaching, cognitive and social presence on cognitive load. *The Internet and Higher Education, 31*, 11–19. http://doi.org/10.1016/j.iheduc.2016.05.003

Kozan, K., Erçetin, G., & Richardson, J. C. (2015). Input modality and working memory: Effects on second language text comprehension in a multimedia learning environment. *System, 55*, 63–73. http://dx.doi.org/10.1016/j.system.2015.09.001

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*, 229–270. http://dx.doi.org/10.1111/j.1467-9922.2007.00408.x

Li, S. (2017). Cognitive differences and ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 396-417). New York: Routledge.

Li. S., & Fu, M. (2018). Strategic and unpressured within-task planning and their associations with working memory. *Language Teaching Research, 22*(2), 230–253. http://doi: 10.1177/1362168816684367

Lindstromberg. S. (2020). Intentional L2 vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 240-254). New York: Routledge.

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60*(3), 501–533. http://doi:10.1111/j.1467-9922.2010.00565.x

Mackey, A., Philip, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–209). Amsterdam: John Benjamins.

Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning, 62*, 704–740. http://doi.org/10.1111/j.1467-9922.2011.00649.x

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*, 179-413. http://doi:10.1017/S0272263112000125

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

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

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. http://doi:10.1207/S15326985EP3801_6

Miller, Z. F., Fox, J. K., Moser J. S., & Godfroid, A. (2018): Playing with fire: Effects of negative mood induction and working memory on vocabulary acquisition, *Cognition and Emotion*, *32*(5), 1105–1113. http://dx.doi.org/10.1080/02699931.2017.1362374

Nation, P. (2001). *Learning vocabulary in another language*. Harlow: Longman.

Nowbakht, M., & Fazilatfar, M. A. (2019). The effects of working memory, intelligence and personality on English learners' speaking ability. *The Journal of Asia TEFL, 16*(3), 817-832. http://dx.doi.org/10.18823/asiatefl.2019.16.3.4.817

Paribakht, S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 147–200). Norwood, NJ: Ablex.

Phillips, D. (2015). *Longman preparation course for the TOEFL iBT test*. White Plains, NY: Pearson Education.

Plass, J. L., & Jones, L. C. (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.

Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. *Language Learning & Technology*, *23*(2), 105–124. http://doi.org/10125/44685

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning, 62*, 93–132. http://doi.org/10.1111/j.1467-9922.2011.00690.x

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–27). Clevedon, UK: Multilingual Matters. http://doi: 10.21832/9781853599286-004

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* New York: Palgrave Macmillan.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction, 20*(2), 100–110. http://doi:10.1016/j.learninstruc.2009.02.011

Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review, 23*(3), 389–411. http://doi:10.1007/s10648-011-9168-5

Skehan, P. (2012). *Researching tasks: Performance, assessment, pedagogy*. Shanghai: Shanghai Foreign Language Education Press/Amsterdam: De Gruyter.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review, 64*(3), 429–458. http://doi:10.3138/cmlr.64.3.429

Speciale, G., Ellis, N. C., & Bywater, T. (2004) Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics 25*, 293–321. https://doi.org/10.1017/S0142716404001146

Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review, 22*, 123–138. https://doi.org/10.1007/s10648-010-9128-5

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

Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71–81. https://doi.org/10.1348/000709904322848824

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 171–195). Oxford: Oxford University Press.

Türk, E., & Erçetin, G. (2012). Effects of interactive versus simultaneous display of multimedia glosses on L2 reading comprehension and incidental vocabulary learning. *Computer Assisted Language Learning, 27*(1), 1–25. http://dx.doi.org/10.1080/09588221.2012.692384

Van den Noort, M. W. M. L., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*(4), *289-296.* http://doi:10.1027/1016-9040.11.4.289

Wang, S., & Lee, C. (2020). Exploring the effectiveness of hypermedia glosses for second language reading. *The Journal of Asia TEFL, 17*(4), 1284-1293. http://dx.doi.org/10.18823/asiatefl.2020.17.4.8.1284

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.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). New York: Routledge.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly, 43*(1), 55–77. https://doi.org/10.1002/j.1545-7249.2009.tb00227.x

Wen, Z. (2016). *Working memory and second language learning: Towards an integrated approach.* Bristol: U. K: Multilingual Matters.

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2009). Observational learning from animated models: Effects of modality and reflection on transfer. *Contemporary Educational Psychology, 34*, 1–8. http://doi:10.1016/j.cedpsych.2008.03.001

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System, 70*, 38–49. http://doi: 10.1016/j. system.2017.09.012

Yoshii, M. (2013). Effects of gloss types on vocabulary learning through reading: Comparison of singlet and multiple-choice gloss types. *CALICO Journal, 30*, 203-229. Retrieved February 20, 2021, from https://www.jstor.org/stable/calicojournal.30.203

Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, *24*(1), 39–58. http://dx.doi.org/10.1080/09588221.2010.523285

Appendix A

Text-plus-picture Glosses

Appendix B

Text-only Glosses