





# Task-induced involvement load and working memory: Effects on active and passive vocabulary knowledge of EFL learners in a multimedia learning environment<sup>+</sup>

Ali Akbar Ansarin<sup>a</sup> , Salva Kazemipour Khabbazi<sup>a\*</sup> 

<sup>a</sup> Department of English Language and Literature, University of Tabriz, Tabriz, Iran

Received 15 October 2020

Received in revised form 17 February 2021

Accepted 26 February 2021

## APA Citation:

Ansarin, A. A. & Kazemipour Khabbazi, S. (2021). Task-induced involvement load and working memory: Effects on active and passive vocabulary knowledge of EFL learners in a multimedia learning environment. *Eurasian Journal of Applied Linguistics*, 7(1), 277–302.

Doi: 10.32601/ejal.911288

## Abstract

This study aims at comparing the effect of single- and dual-annotation modes, working memory (WM), and three tasks of different involvement loads (sentence writing, close deletion, and paragraph writing) on passive and active vocabulary development of proficient English as a foreign language (EFL) learners. While listening to an expository text, participants ( $n = 204$ ) were presented with annotations for twenty lexical items. Paribakht and Wesche's (1997) Vocabulary Knowledge Scale was employed as the pre and posttests and via a listening span task, the learners' WM capacities were measured. The results pointed to the efficacy of dual-mode over single-mode annotations concerning both passive and active vocabulary learning. The participants with higher WM capacities significantly performed better regarding active vocabulary learning for both textually and visually annotated words, thus no interaction between WM and input mode was found. The sentence writing task produced the highest scores in the vocabulary tests, whereas the cloze deletion and paragraph writing tasks did not significantly differ. Hence, the results only partially confirmed the Involvement Load Hypothesis. The current study corroborates other studies verifying the modality effect on vocabulary learning and suggests that both WM and task type are significant factors in EFL vocabulary learning.

© 2021 EJAL & the Authors. Published by *Eurasian Journal of Applied Linguistics (EJAL)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Annotations; working memory; involvement load; passive vocabulary development; active vocabulary development

## 1. Introduction

<sup>+</sup> This study is based on the dissertation titled “xxx” that is submitted to University of Tabriz by “xxx”.

\* Corresponding author. Tel.: +98-914-107-2417

E-mail address: [kazemipour\\_s@tabrizu.ac.ir](mailto:kazemipour_s@tabrizu.ac.ir)

<http://dx.doi.org/10.32601/ejal.911288>

An increasing number of studies are indicative of the importance of explicit, and elaborate word-focused activities to successfully improve second language (L2) vocabulary learning (Laufer, 2005; Schmitt, 2008). Laufer and Hulstijn (2001) proposed the Involvement Load Hypothesis (ILH), according to which tasks involving higher cognitive load lead to more vocabulary acquisition than those with lower involvement loads. This hypothesis, however, has only been investigated with regard to single tasks or in post-reading activities (Kim, 2011; Laufer, 2003). On the other hand, Cognitive Theory of Multimedia Learning (CTML) and Cognitive Load Theory (CLT) have been emphasizing the effect of multimedia learning on promoting vocabulary development (Mayer, 2009; Sweller, 2010) on the assumption that written information together with visual aid will overload the WM by competing for its visual resources. Consequently, presenting materials in an audiovisual manner is proposed. One of the means to realize this aim is the use of annotations to represent new lexical items in the course of reading and listening activities (Nation, 2013). The impact of type (picture, video, or text) and mode (single, dual or multimodal) of annotations along with the system of instruction (system- or self-paced) is open to dispute (Boers, Warren, He, & Deconinck, 2017). Nevertheless, the modality effect and the use of annotations have mostly been explored on text comprehension or vocabulary learning through reading, with few studies on lexical retention through listening (Çakmak & Erçetin, 2017).

Aside from the unsettled issues on annotations, the role of different components of WM on mediating the modality effect and tasks with a variety of involvement loads has not been thoroughly investigated. In addition to the relationship between WM and the modality effect, there is an association between WM and the “evaluation” component of the ILH (Martínez-Fernández, 2008). Thus, this paper first aimed to look more thoroughly into the modality effect and the use of different annotation modes to improve both passive and active vocabulary development. Second, it sought to inspect if WM, the executive WM in particular, and tasks of different involvement loads mediate lexical retention through listening to an audio text and in post-listening activities. In order to fill the abovementioned gaps, this study seeks to address the following research questions:

1. Are different modes of input (i.e. text-only versus text-plus-picture annotations) equally effective on passive and active vocabulary development of EFL learners?
2. Do differences in working memory capacity affect EFL learners’ gain scores of passive and active vocabulary knowledge?
3. Do tasks (i.e. sentence writing, cloze deletion, and paragraph writing) with different involvement loads lead to differential gains of passive and active word knowledge for EFL learners?
4. Are there any interaction effects of working memory capacity, task types of varying involvement loads, and different modes of input (text-only versus text-plus-picture annotations) on EFL learners’ passive and active knowledge of words?

## **2. Literature Review**

### *2.1. Multimedia annotating and vocabulary knowledge*

Multimedia instruction is a contributing factor in fostering learning process (Mayer, 2009). According to the modality effect, presenting texts in spoken format is more effective than written texts when accompanied by pictures, illustrations, diagrams or charts (Mayer, & Moreno, 2003). The theoretical rationale of the modality effect is based upon CTML which assumes that learning takes place through WM. Since WM has limited capacity, audiovisual materials have been superior to visual-only ones (Mayer, 2009; 2014). It is due to the fact that channels for visual and auditory input are separate in WM, hence less load on WM via audiovisual instruction. Most of the research has supported the superiority of audiovisual presentation (Kozan, Erçetin, & Richardson, 2015). Some studies, however, have found no modality effect (De Westelinck, Valcke, De Craene, & Kirschner, 2005) or reverse modality effect (Tabbers, Martens, & van Merriënboer, 2004) which occurs when learning improves when the learners are exposed to a written text plus pictures (i.e. a visual presentation). It has been assumed that self-paced systems plus complex written texts cause the disappearance of the modality effect (Tabbers & de Koeijer, 2010).

As one of the means to facilitate comprehension and vocabulary learning, annotations offer external aids by providing translations or meanings of unknown words in a text. They can contribute to the modality effect if they contain pictures, videos or other types of visual illustrations. Although the use of annotations has proved to be beneficial (Jones & Plass, 2002; Yun, 2011), type of annotations advantageous for vocabulary learning has been open to dispute. The difference between the effect of single-mode and dual-mode annotations is not clear. Some studies have reported on the advantage of dual-mode annotations (Lin & Tseng, 2012), some others, however, have found no significant benefit of duality in annotations on vocabulary learning (Boers, Warren, He, et al., 2017; Çakmak & Erçetin, 2017). The discrepancies in the results call for further research investigating the effect of multimedia annotations on vocabulary learning.

Moreover, the majority of the experiments has been on the effect of multimedia annotations in a reading context, with very few of them examining the effect through a listening text (Cottam, 2011; Çakmak & Erçetin, 2017), the results of which have been inconsistent. The need to further explore the effect of multimedia annotations on L2 vocabulary learning through listening seems apparent. As a result, one of the questions addressed in this paper was the effect of single- (text-only) and dual-mode (text-plus-picture) annotations on passive and active vocabulary knowledge of proficient English learners.

### *2.2. Working memory in EFL vocabulary learning*

Another variable considered in the study was WM. WM is a system responsible for both the storage and processing of information in the face of cognitive tasks such as comprehension, reasoning, or some distractions (Baddeley, 2015). The role of WM in L2 processing and acquisition along with the increase of L2 studies on specifying its role in language acquisition has been asserted. One of the central components of WM is short-term memory (or phonological short-term memory, i.e. PSTM) which is responsible for the temporary storage of information. The other is called the Central Executive (CE) or executive WM which is concerned with both the storage and processing of information (Baddeley, 2007).

Despite the overall well-established role of WM in L2 development (Juffs & Harrington, 2011; Williams, 2011), whether PTSM or CE component of WM is responsible for L2 vocabulary learning is still inconclusive. While most of the research has reported on the predictive power of PTSM in L2 vocabulary acquisition in earlier stages and for less proficient learners (French & O'Brien; 2008), other scholars have pointed out the correlation between executive WM and the amount of vocabulary learned in an L2 context (Linck, Osthus, Koeth, & Bunting, 2014) and for more proficient learners (Yang, Shintani, Li, & Zhang, 2017). Yet, the statement that more proficient learners tend to draw on executive WM more than PTSM needs more evidence and confirmation through further research.

Furthermore, the amount of research on the involvement of WM during multimedia instruction is little with inconsistencies in results. Kozan et al. (2015) aimed to examine the effect of two types of visual-only and audiovisual multimedia presentations and WM on reading and listening comprehension. The results suggested an interaction between the modality effect and WM as there was a higher retention rate for participants with higher WM than those with lower WM. They also pointed out the important role WM plays in listening, which is even more than it is in reading. Similarly, the results in Gyselinck, Jamet, & Dubois (2008) displayed the involvement of verbal WM in multimedia learning for both visual and auditory presentations. A general WMC effect was also found in a multimedia learning environment (Doolittle, Terry, & Mariano; 2009), which was interpreted to be attributed to attentional resources that can be employed by learners to different extents. This is while Sanchez and Wiley (2006) gained confounding results regarding the WM effect on learners' comprehension

In addition to the unresolved role of WM in multimedia learning, the abovementioned studies have all focused on the involvement of WM in text comprehension, either reading or listening. As the effect of WMC on passive and active vocabulary learning through a listening text in a multimedia environment has not yet been explored, one of the goals of the current study was to shed light on this issue.

### *2.3. Task-induced involvement load*

Laufer and Hulstijn (2001) have proposed the ILH, according to which the amount of involvement load of a task is the determiner of the extent to which unknown lexical items are learned. Consisting of three motivational-cognitive constructs; that is, need, search, and evaluation, the ILH can assess the degree to which a task engages learners in learning. As the motivational dimension of the ILH, the need to act in accordance with task requirements (i.e. need) might be moderate (with an index of 1) or strong (with an index of 2). It is moderate when it is extrinsically induced and the task or the teacher requires learners to use particular vocabulary in a sentence or text. It is strong if intrinsically induced and learners feel the need to understand or express a concept using a source such as a dictionary. Two cognitive dimensions, search and evaluation, are mainly conditional on noticing and deliberate attention to the form and meaning of words in a task. Search is present if learners consult an external source (i.e. the teacher or the dictionary) to find the meaning or form of an unfamiliar word to complete a task. In this case, it has an index of 1 but when it is absent, the index is 0. Evaluation involves a comparison of a word with others, one meaning of a word with other meanings of the same word, or an assessment of whether a word appropriately fits a new context or not. Evaluation also has two degrees of prominence. It would be moderate (an index of 1) if the task necessitates a comparison among words. It is strong (an index of 2) when learners assess the appropriateness of words in a new context. The quantifiable nature of the ILH allows researchers to precisely determine the efficiency of different task types in promoting L2 vocabulary acquisition. The involvement index of a task is the sum of the scores for these three dimensions and ranges from 0 to 5. It is hypothesized that the higher the involvement load, the more efficient a task would be for learning unknown vocabulary items (Hulstijn & Laufer, 2001; Kim, 2011).

In their work, Hulstijn and Laufer (2001) provided full and partial support to the hypothesis. The same is true for the follow-up studies whose results were controversial with a small number fully supporting the ILH (Bao, 2015; Kim, 2011; Yang et al. 2017). These studies have also improved the hypothesis by adding other contributing factors such as time on task, proficiency level of learners, exposure frequency, and vocabulary knowledge tests used.

Taken together, all these studies tested the ILH in a reading context (Keating, 2008; Kim, 2011; Yang et al. 2017) or a sentence-context design (Bao, 2015; Pichette, de Serres, & Lafontaine, 2012). To the best of our knowledge, no research has implemented the ILH on post-listening activities on passive and active L2 vocabulary knowledge. Therefore, four purposes of the current study are: 1) to expand on the research into multimedia instruction by examining the effect of single- and dual-mode annotations on EFL passive and active vocabulary knowledge; 2) to explore WM effect on passive and active vocabulary learning through a listening text with text-only and text-plus-picture annotations; 3) to broaden the research on the ILH to the effect of post-listening tasks with different loads on passive and active vocabulary learning; and 4) to investigate the interaction of multimedia annotations, tasks with different involvement loads, and WM on EFL vocabulary learning.

### 3. Method

#### 3.1. Participants

The participants were selected from three public universities in East-Azerbaijan, Iran. A total of 204 students (146 females and 58 males;  $M = 20.47$  years) majoring in English language and literature in eight intact classes participated in the study. In order to be enrolled in the Department of English, they had to take the special English language test of the National University Entrance Examination consisting of vocabulary, grammar, reading comprehension, language function, and cloze modules at an advanced level. As their scores were beyond 75 percent in this test, they can be considered as B2 level or independent users according to CEFR. Initially, there were 245 participants, some of whom dropped out of the study ( $n = 18$ ) and others were excluded from the analysis as their z-scores were beyond the normal range ( $n = 23$ ).

#### 3.2. Research Design

The study followed a pretest, treatment and posttest design with the vocabulary pretest in week 1, the working memory test in week 2, and the treatment with the immediate vocabulary posttest in week 3. Both passive and active vocabulary knowledge of EFL learners were measured in the pre and posttests. In the treatment session, the vocabulary items were presented simultaneously with the listening text through single- (i.e. text-only) and dual-mode (i.e. text-plus-picture) annotations. After the presentation, the participants were asked to answer five comprehension questions about the listening text. Then, the participants, randomly divided into three groups, completed tasks with different involvement loads followed by the vocabulary posttest. Hence, the study adopted a partially within-subjects and partially between-subjects design. Through all stages, the participants were not allowed to use any dictionaries or electronic devices, nor could they discuss the words with each other.

#### 3.3. Instructional materials

The listening text [mp3 039-040], selected from the Third Edition of Longman Preparation Course for the TOEFL iBT Test (Phillips, 2015, p. 172), is a lecture about opossums and their defense mechanisms. At first, twenty-five vocabulary items were selected. However, after a pilot test with similar participants, five of the items were eliminated due to the familiarity of the learners with those five items. The twenty selected items, with which the participants were assumed to be unfamiliar, were afterward divided into two groups to be presented through single-mode (text-only, i.e. Persian translation of the words) or dual-mode (text-plus-picture, i.e. Persian translation together with a picture representing the meaning of the word) annotations. Ten words of the text-only annotations 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 vocabulary items for the text-plus-picture annotations were: cling (v.), emit (v.), marsupial (n.), opossum (n.), outrun (v.), play possum (phr.),

pouch (n.), racket (n.), stench (n.), venom (n.). The criterion for categorizing the words into text-only and text-plus-picture was the possibility of lucidly depicting them through pictures. Therefore, abstract words were selected for text-only annotations. The text-plus-picture glosses are included in the appendix. Nevertheless, all of these words were checked for their frequency and they were of low frequency as they are off the 2000 frequency word lists, Academic Word list (AWL) and University Word lists (UWL).

The participants were presented with the expository listening text as well as visual slides in a video file through video projectors. The slides contained the target words with either their Persian translations (L1) or their translations and corresponding pictures. They appeared on the screen the instant that the lecturer in the listening text mentioned them. Opting for a system-paced presentation condition, each slide remained on the screen for 4 seconds intending to measure the amount of vocabulary acquisition through listening to a text accompanied by visual aids.

### *3.4. Post-listening activities and task involvement load*

After listening to the text and answering five multiple-choice listening comprehension questions, the participants completed one of the three tasks; that is, sentence writing (n = 67), cloze deletion (n = 68), and paragraph writing (n = 69). The allowed time was 30 minutes for all the tasks. The participants, however, were asked to write down the amount of time they spent on the tasks. The participants in the Sentence writing group needed to write an authentic sentence of more than seven words for each of the twenty words. They were provided with the L1 translation of the words to refer to during the task if needed. The involvement load index of this task is 3, no search, a moderate need, and a strong evaluation (Kim, 2011). The Cloze deletion group read a 318-word text about opossum and filled out the blanks with the appropriate words from a list including the twenty words and four distractors together with their L1 meanings. The involvement index of this task is 2, as there is no search, a moderate need, and a moderate evaluation. The last task required the Paragraph writing group to write a paragraph with a maximum of 100 words summarizing the listening text. As the task did not require the participants to employ the selected words in their text, nor did it include any explicit focus on the words, the involvement load index was 0 (no need, no search, and no evaluation). Subsequent analysis of the texts indicated that the participants in this group adopted some type of avoidance strategy in order not to use the target words in their writings. Except for the word “opossum”, which was used by some of the participants, the rest of the target words were ignored in the paragraphs.

### *3.5. Testing instruments*

Immediately after the listening text and visual presentation, the participants were required to answer five multiple-choice comprehension questions. It was aimed to ascertain that besides the target words the participants had paid attention to the

content of the text as well. It is asserted that questions appraising local comprehension pertain to WMC more than global comprehension questions (Brunfaut and Révész, 2014). As a result, all five questions included detail information presented in the listening test.

### 3.5.1. Vocabulary test

There is a continuum along which a learner's knowledge of words is likely to move with some aspects learned before others (Schmitt, 2010). It is also mentioned that a large passive vocabulary knowledge rarely leads to efficient language use and the ability to use the words actively requires deep knowledge of vocabulary items (Anderson & Freebody 1981, as cited in Schmitt, 2010). This suggests that the link created between the meaning and the form of vocabulary items, merely adequate for recognition, is insufficient for active use of language. In line with this notion and the aim to measure both passive and active vocabulary knowledge, Paribakht and Wesche's (1997) Vocabulary Knowledge Scale (VKS) was used as both the pretest and posttest. This scale has been used frequently in studies on vocabulary development (File & Adams, 2010; Kim, 2011; Yang et al., 2017). It has a five-point scale of vocabulary knowledge through which self-perceived vocabulary knowledge and actual performance are scored.

As differences in the effect of the treatment on passive and active knowledge were of importance in the present study, the participants were given two sets of scores. One of the scores was for the passive knowledge of the participants (based on 1-4 categories of the test), the other one was given for their active knowledge of the vocabulary items (based on category 5). The first and second categories of passive knowledge, 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", just required a tick in the box. The third and the fourth categories were 3: "I have seen this word before, and I think it means", 4: "I know this word it means", that required a synonym or the Persian translation of the words. The active knowledge score was given based on category 5: "I can use this word in a sentence", for which the participants were asked to write an original sentence. Ticking category 1 meant no knowledge of the item, hence no score was given. Putting a checkmark in category two meant the word just seemed familiar, so a score of 1 was awarded. A score of 2 and a score of 3 were awarded for categories 3 and 4 respectively. If the synonym or the meaning provided in category 3 or 4 was incorrect, 2 points were given. It is worth mentioning that VKS allows the researchers to measure the degree to which the participants are certain of the meanings they provide (i.e. two points when they are not positive about the meaning and three points when they are confident). Consequently, it was possible to measure even minor changes in the lexical knowledge of the participants as a result of the treatment. In the active category of 5, no sentence received no scores, a semantically correct sentence was awarded 4 points, and a semantically as well as grammatically correct sentence received a score of 5. It is worth mentioning that simple sentences indicating inadequate or imperfect knowledge of the words did not receive any scores given the



relative proficiency of the participants. The gain scores were calculated separately for passive and active knowledge before and after the treatment through the pre and posttests (i.e. post - pretest score). Therefore, four scores were given to each participant; that is, one passive and one active score for verbally annotated words together with one passive and one active score for visually annotated words.

Both the pre and the posttests were in English with a short Persian introduction given by the researcher to ensure that the procedure was understood. In order to avoid rote learning and carry-over effect from the previous sessions, the 20 vocabulary items and 4 distractors appeared in a random order in the pretest, posttest and also in the sentence writing task. The participants were informed of the forthcoming posttest. It should be mentioned that one of the researchers together with an independent researcher scored the VKS tests (the pre and posttests) for all of the 204 participants. Interrater reliability was high (95%). Nevertheless, all the disagreements were deliberated on until agreement was reached.

### *3.5.2. Working memory test*

The listening span test employed in the current study was the spoken version adopted from the reading span test of Daneman and Carpenter (1980). This test demands both the judgment of sentence plausibility and the memorization of the last word of the sentences; that is, process and storage (Mackey, Adams, Stafford, & Winke, 2010). As this test taps into language related aspects of WMC, it is related to the purpose of the study.

The test included 60 English sentences which were randomly divided into 15 sets consisting of 2-6 sentences. To prevent forming meaningful associations among words in one set which would make their memorization easier, none of the final words were semantically related. In addition, some sentences with difficult final words were replaced by easier sentences to exclude listening skills as a contributing factor in the performance on the test. After instructions in L1 and three practice sets, the audio test was played for the participants and they had to listen to sets of sentences making semantic judgments plus memorizing the last word of each sentence. There was a 2-second pause between sentences within a set for the participants to put a mark next to the number of the sentence. A tick would mean the sentence was semantically plausible and an x would mean it was not. At the end of each set, they had to recall and write down the final words of the sentences. All sentences were numbered in the audio file and the answer sheets. The participants were given 5 seconds to recall each final word. Therefore, if the set contained 4 sentences, the allocated time to write down the final words of the set was 20 seconds. The order in which the final words were written in the answer sheet did not have to be the same as the audio file (Chen, 2013). When recalling time was over, a beep sound announced the beginning of a new set.

Partial-credit load scoring procedure (Conway et al., 2005) was chosen to calculate the scores of the storage aspect of the test. Accordingly, for the correctly recalled final words 1 score was awarded, irrespective of their position in the set or the number of

sentences in each set. Incorrectly recalled items were not given any points. The maximum possible score was 60. As for the process aspect of the test (i.e. semantic judgment), a high correlation is shown in most studies between the performance on the storage and processing tasks (Kane et al., 2004). Additionally, it is asserted that since being attentive to the semantic plausibility of the sentences is emphasized, the accuracy of judgments is usually close to the ceiling and there is no evidence of a trade-off between the two tasks (Conway et al., 2005). Consequently, the data for the participants whose judgment scores were beyond 85% of accuracy were kept in the analysis (all the participants had judgment scores beyond 85% accuracy).

### 3.6. Data analysis

Data analysis in this study consisted of several steps. Firstly, the participants' gain scores on the VKS pre and posttests were calculated considering the two input modes (text-only and text-plus-picture annotations) and passive and active knowledge. That meant there were four scores for each participant: passive knowledge of the words presented in the single- and the dual-mode annotations; active knowledge of the words presented in the single- and dual-mode annotations. A paired-sample t-test was conducted to answer the first research question; that is, to compare the effect of single- and dual-mode annotations on passive and active word acquisition. To investigate the effect of WM on vocabulary acquisition, the scores of the participants were divided into two high and low groups based on the median split method. The use of the median split method has been validated in other research (e.g. Iacobucci, Posavac, Kardes, Schneider, & Popovich, 2015). A two-way MANOVA was conducted on the abovementioned four scores with WM and task type as between-subject variables to explore their main effects (research questions 2 and 3) together with the interaction effect of WM, and task type on passive and active vocabulary knowledge in two annotation modes (research question 4). Time on task was inserted into the design as a covariate to control the effect of the differences in time spent on the tasks. In order to find out about any relationships between listening comprehension questions and WMC, a simple linear regression was conducted to analyze to what extent WMC mediated learners' listening comprehension. All the analyses were executed using SPSS version 23.

## 4. Results

Table 1 indicates the descriptive results for the listening span task and the gain scores of the participants in the two VSK tests (pre and posttests).

### 4.1. Multimedia annotations and EFL vocabulary knowledge

The first research question concerned how single- and dual-annotation modes influence EFL passive and active vocabulary knowledge. For this purpose, a paired-sample t-test was conducted to determine whether there was a statistically significant mean difference between the gain passive and active scores for words under text-only

and text-plus-picture annotations. The results, presented in Table 2, indicate that the effect of the two annotations on both passive and active vocabulary learning was significantly different.

There was a significant difference between the passive gain scores for words under text-only annotation ( $M = 11.92, SD = 5.24$ ), and under text-plus-picture annotation ( $M = 15.50, SD = 5.09$ );  $t(203) = -10.98, p = .0001$ . Moreover, with Cohen's effect size value of  $d = 0.69$ , a moderate to high practical significance can be observed. Similarly, active gain scores of vocabulary in text-only ( $M = 22.08, SD = 11.76$ ), and text-plus-picture annotations ( $M = 28.29, SD = 11.44$ ), were found to be significantly different;  $t(203) = -8.65, p = .0001$ , with a moderate effect size (Cohen's  $d = .52$ ). Hence, both null hypotheses that single- and dual-annotation modes do not have an effect on passive and active vocabulary knowledge were rejected. The results supported the multimedia effect by demonstrating multimedia annotation effect on both passive and active vocabulary knowledge, with dual-mode annotations (i.e. text-plus-picture) leading to more vocabulary learning.

Table 1. Descriptive statistics for VKS test and WM task

	Passive		Active		WMC	
	Knowledge		Knowledge		M	SD
	M	SD	M	SD		
Single-mode annotation	11.92	5.246	22.08	11.767		
Dual-mode annotation	15.50	5.091	28.29	11.443	31.14	7.922

n: 204

Table 2. T-test results of the differences between two annotation modes

	n	Mean	SD	t	df	Sig. (2-tailed)	d
Pair 1 Passive, Mode 1 & 2	204	-3.588	4.665	-10.987	203	.000	.692
Pair 2 Active, Mode 1 & 2	204	-6.216	10.259	-8.653	203	.000	.528

$p < .05$

#### 4.2. WM, multimedia annotations, and vocabulary development

To answer the second, third, and fourth question, a two-way MANOVA was conducted. To check the assumptions of a two-way MANOVA, three main tests, that is Levene's test of equality of variances, Box's M for equivalence of covariance matrices,

and Bartlett's test of sphericity, were used. Levene's test of equality of variances was insignificant for all of the four sets of VKS scores (passive word knowledge of text-only annotations:  $F = 2.356$ ,  $p = .049$  and text-plus-picture annotations:  $F = .787$ ,  $p = .560$ , active word knowledge under text-only annotations:  $F = .357$ ,  $p = .877$ , and text-plus-picture annotations:  $F = .288$ ,  $p = .919$ ) indicating homogeneity of variance. Box's M [ $F = 1.211$ ,  $p = .146 > \alpha = .01$ ], was insignificant suggesting that the covariance matrices was equal for the groups. Finally, as Bartlett's test of sphericity was significant ( $\chi^2 = 476.082$ ,  $p = .0001 < \alpha = .01$ ), it was appropriate to use the factor analytic model on this set of data. Table 3 summarizes the main effect of the variables together with the interaction effect of WM and task type. The results for the main effect of WM will be discussed in this section.

Table 3. Main effect and interaction of the three variables

Variables	Two-way MANOVA									
	Main Effect					Interaction Effect				
	V	df	F	Sig.	$\eta^2$	V	df	F	Sig.	$\eta^2$
Working memory	0.170	1;197	9.960	0.0001*	0.170	-	-	-	-	-
Task type	0.235	2;197	6.487	0.0001*	0.117	-	-	-	-	-
Time On Task	0.019	1;197	0.927	0.450	0.026					
WMC *Task type						0.052	2;197	1.297	0.243	0.26

The second research question asked whether the WM played a role in the passive and active vocabulary development of advanced EFL learners. The Pillai's trace was indicative of the overall significant main effect of WM on the participants' vocabulary development with a large effect size [ $F(1, 198) = 9.960$ ,  $p = .0001$ , partial  $\eta^2 = 0.170$ ]. The effect of WM on the four sets of scores is as follows. The F ratio of passive word knowledge under single-mode glosses was  $F(1,198) = .288$ ,  $p = .592$ , partial  $\eta^2 = 0.001$ , showing an insignificant difference between high-WM ( $M = 12.28$ ,  $SD = 5.53$ ) and low-WM ( $M = 11.60$ ,  $SD = 5.20$ ) groups. Similarly, the main effect for passive vocabulary development under dual-mode glosses yielded an F ratio of  $F = 1.266$ ,  $p = .262$ , partial  $\eta^2 = 0.006$ , also failing to indicate a significant difference between high-WM ( $M = 16.03$ ,  $SD = 5.33$ ) and low-WM ( $M = 15.05$ ,  $SD = 4.98$ ) groups.

On the other hand the results for active aspect of word knowledge were rather different. The F ratio for active vocabulary development for words presented through single gloss mode was  $F(1, 198) = 17.122$ ,  $p = .0001$ , partial  $\eta^2 = 0.080$ , illustrating a significant difference between high-WM ( $M = 25.49$ ,  $SD = 10.73$ ) and low-WM ( $M = 19.10$ ,  $SD = 11.86$ ) groups with a medium effect size. Likewise, the effect of WMC on active vocabulary development under dual gloss mode was found to be significant with an F ratio of  $F(1,198) = 24.45$ ,  $p = .0001$ , partial  $\eta^2 = 0.110$ , which was indicative of the difference between the performance of high-WM group ( $M = 32.28$ ,  $SD = 10.36$ ) and low-WM group ( $M = 28.82$ ,  $SD = 11.24$ ) with a large effect size. The null

hypothesis stating that WMC did not play a role in vocabulary development was rejected. However, this rejection was exclusive to active vocabulary knowledge.

#### 4.3. Post-listening activities and vocabulary development

The third research question asked if tasks of different involvement loads affect passive and active vocabulary development. Table 4 summarizes the descriptive statistics for the performance of the three task type groups on VKS. According to the Pillai's trace, the task type indicated a significant overall main effect on vocabulary development at the  $p < .05$  level for paragraph writing, cloze deletion, and sentence writing conditions with a large effect size [ $F(2, 198) = 6.487, p = .0001, \text{partial } \eta^2 = .117$ ]. Results pointed to the significant differences that all the four categories of vocabulary development have among the three task conditions: (a) passive knowledge in text-only annotation mode: [ $F(2, 198) = 12.917, p = .0001, \text{partial } \eta^2 = 0.116$ ], (b) passive knowledge in text-plus-picture annotation condition: [ $F(2, 198) = 5.552, p = .0005, \text{partial } \eta^2 = 0.053$ ], (c) active knowledge in text-only: [ $F(2, 198) = 26.045, p = .0001, \text{partial } \eta^2 = 0.209$ ], and (d) in text-plus-picture conditions: [ $F(2, 198) = 8.950, p = .0001, \text{partial } \eta^2 = 0.083$ ].

Table 4. Descriptive statistics for the performance on VKS by different task type groups

	Passive			Passive		Active		Active	
	Single-mode			Dual-mode		Single-mode		Dual-mode	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Paragraph	69	10.16	3.79	14.26	4.81	16.25	9.59	23.09	10.51
Cloze	68	10.69	5.25	14.87	5.31	19.24	10.55	27.54	11.38
Sentence	67	14.97	5.24	17.43	4.62	30.97	9.68	34.42	9.49

Post hoc comparisons using the Tukey HSD test indicated that the mean score of the Sentence writing group was significantly higher than that of the Cloze deletion and Paragraph writing groups in passive vocabulary development under both multimedia annotations. The same is true for active vocabulary development under text-only annotation mode, though the effect size of the difference between the Sentence writing and Cloze deletion groups was low ( $d = .25$ ). This is while the Cloze deletion group's performance did not significantly differ from the Paragraph writing group. Regarding active vocabulary development under dual-mode annotations, all three groups performed variously, with the Sentence writing group significantly outperforming the other groups (very high and high effect sizes), and the Cloze deletion group performing better than the Paragraph writing group (a medium effect size). The results of Tukey HSD along with the effect size (Cohen's  $d$ ) are summarized in Table 5.

The results also pointed to the insignificant main effect of time on task as the covariate. The Pillai's trace showed an insignificant main effect of time on task on passive and active vocabulary development of the participants [ $F(2, 198) = 0.927, p = 0.450, \text{partial } \eta^2 = 0.019$ ]. Therefore, it can be concluded that time on task did not affect the scores of the participants in different task groups. Overall, the results only partially supported the ILH, as the performance of the Cloze deletion and Paragraph writing groups, despite the tasks' various involvement indices, were not significantly different except for the active vocabulary development for the words which were presented through text-plus-picture annotations.

Table 5. Levels of significance and effect sizes between three task type groups

	Passive		Passive		Active		Active	
	Single mode		Dual mode		Single mode		Dual mode	
	Sig.	d	Sig.	d	Sig.	d	Sig.	d
Sentence-paragraph writing	.000	1.05	.001	.67	.000	.54	.000	1.31
Sentence writing-Cloze	.000	.81	.008	.51	.000	.25	.001	.65
Paragraph writing-Cloze	.794		.751		.187		.001	.42

$p < .05$

#### 4.4. The interaction between WM and task type on vocabulary development

It has been observed that WM and task type, although dissimilarly, have independently contributed to vocabulary development. In order to answer the forth research question to examine the interaction effect of WM (high versus low) and task type (Paragraph writing, Cloze deletion, and Sentence writing groups) on passive and active vocabulary development under single- and dual-annotation modes, the interaction effect in the two-way MANOVA test is analyzed and interpreted. Table 6 and 7 demonstrate the descriptive statistics for the performance of high-WM and low-WM groups on the VKS tests based on the tasks they completed. As Pillai's trace demonstrated, there is no significant interaction effect between the two variables [ $F(2, 198) = 1.297, p = 0.243, \text{partial } \eta^2 = 0.026$ ] on the vocabulary development of the participants.

Except for active vocabulary knowledge in text-only condition which showed a significant interaction effect of WM and task type [ $F(2, 198) = 4.773, p = .009, \text{partial } \eta^2 = .046$ ], in the other three conditions; that is, passive vocabulary knowledge in text-only [ $F(2, 198) = 2.170, p = .117, \text{partial } \eta^2 = .021$ ], text-plus-picture annotation modes [ $F(2, 198) = .496, p = .610, \text{partial } \eta^2 = .005$ ], and active knowledge in text-plus-picture annotation condition [ $F(2, 198) = 1.63, p = .204, \text{partial } \eta^2 = .16$ ], the interaction effect of WM and task type on vocabulary development was insignificant.

The pairwise comparison between WM and task type indicated that in active vocabulary knowledge in text-only condition, the performance of the high-WM

participants in the VKS was significantly different between the three task type groups. In other words, the Sentence writing group ( $M = 31.46$ ,  $SD = 9.86$ ) outperformed both the Cloze deletion group ( $M = 25.52$ ,  $SD = 8.77$ ),  $p = .013$ ,  $d = .63$ , and the Paragraph writing group ( $M = 14.21$ ,  $SD = 9.28$ ),  $p = .0001$ ,  $d = 1.35$ , along with the better performance of the Cloze deletion group than the Paragraph writing group,  $p = .006$ ,  $d = .77$ . However, the low-WM participants performed similarly in the Paragraph writing ( $M = 14.21$ ,  $SD = 9.28$ ) and Cloze deletion ( $M = 14.56$ ,  $SD = 9.34$ ) groups,  $p = .869$ , while the Sentence writing group ( $M = 30.44$ ,  $SD = 9.60$ ) significantly outperformed the Paragraph writing and Cloze deletion groups,  $p = .0001$ , with very large effect sizes (1.71 and 1.67 respectively). These results suggest that the effect of tasks with various involvement loads was different for the participants with higher WMC, whereas the effect of the paragraph writing and cloze deletion tasks was identical for those with lower WMC.

Table 6. Descriptive statistics for the performance of three high-WM task type groups

	Passive Single-mode		Passive Dual-mode		Active Single-mode		Active Dual-mode	
	M	SD	M	SD	M	SD	M	SD
	Paragraph	10.19	3.64	14.77	5.03	18.74	9.59	26.94
Cloze	11.97	5.67	15.76	5.85	25.52	8.77	33.24	9.95
Sentence	14.40	5.33	17.37	4.97	31.46	9.86	36.23	8.87

Table 7. Descriptive statistics for the performance of three low-WM task type groups

	Passive Single-mode		Passive Dual-mode		Active Single-mode		Active Dual-mode	
	M	SD	M	SD	M	SD	M	SD
	Paragraph	10.13	3.96	13.84	4.65	14.21	9.28	19.95
Cloze	9.74	4.77	14.21	4.83	14.56	9.34	23.31	10.60
Sentence	15.59	5.53	17.50	4.28	30.44	9.60	32.44	9.90

#### 4.5. Working memory capacity and listening comprehension questions

After the listening text plus the annotations were presented, all participants answered five comprehension questions. The comprehension questions served the purpose of making sure that they had not only paid attention to the annotations at the expense of content comprehension. The relationship between vocabulary development and WMC was explored in the previous sessions, accordingly it was sensible to investigate the effect of WMC on listening comprehension as well. Subsequently, we conducted a simple linear regression to analyze to what extent WMC mediated learners' listening comprehension. The results showing  $r = .204$ ,  $n =$

204,  $p = 0.003$ , rejected the null hypothesis which stated that no relationships between WMC and comprehension questions exist. The linear regression equation for WMC and comprehension questions as the independent variable and dependent variable respectively, was  $[F(1,202) = 8.735, p < .001]$  with an  $R^2$  of .041. Participants' predicted comprehension question scores were equal to  $2.044 + .030 \times (\text{WMC scores})$ , so the comprehension questions score increased 0.030 for each point in the WMC test. It was observed that with the  $R^2$  of .041, WMC could weakly predict comprehension question scores of the participants.

## 5. Discussions

### 5.1. Single- versus dual-mode annotations and vocabulary learning

The findings pointing to the superiority of text-plus-picture over text-only annotations in both passive and active vocabulary development provided evidence for the modality effect (Mayer & Moreno, 2003; Mayer, 2009). In conformity with several other studies, dual-mode annotations fostered vocabulary acquisition (Kozan et al. 2015; Lin & Tseng, 2012). One of the reasons mentioned to bring about the modality effect is the use of system-paced instructions in which the learners have no control over the pace of instruction. This notion is in contrast with self-paced instruction which has been mentioned as the root of reverse modality effect (Inan et al., 2013; Tabbers & de Koejier, 2010). As the instruction in the current study was system-paced, it can be asserted that the modality effect is a function of instruction pacing with system-pacing appearing to be more advantageous. In addition, the participants of other studies which found all types of annotations equally effective (Çakmak & Erçetin, 2017) were low proficient learners. It has been stated that proficiency level has an impact on the amount of benefit learners can derive from annotations (Abraham, 2008). Schmitt (2010) also pointed to the threshold level of vocabulary knowledge required before exposure to new words can be of value. We can, thus, attribute the modality effect to the proficiency level of the participants who benefited from dual-mode annotations more.

Besides, the results support CLT which asserts that audiovisual presentations improve learners' performance (Sweller, Ayres, & Kalyuga, 2011). It is also worth mentioning that text-plus-picture annotations have certainly tended to attract more attention due to their nature, therefore more noticing (Nation, 2001). Similar results and related justifications can be found in Jones and Plass (2002) and Boers, Warren, Grimshaw, and Siyanova-Chanturia (2017) which report more retention rates for picture annotated words and point to the amount of engagement picture annotations may cause.

### 5.2. WM effect on vocabulary development through two annotation modes

Passive vocabulary learning of the participants was unaffected by their WM capacities. Although the high WM group performed better than the low WM group, it



fell short of statistical significance. On the other hand, with a medium effect size ( $d = -.56$ ), the high WM group outperformed the low WM group concerning active vocabulary development for textually annotated words. Identically, the high WM group's active knowledge of the words annotated through both verbal and pictorial mode was higher than the low WM group with a medium to high effect size ( $d = -.69$ ). As to the best of our knowledge, no other studies have included active vocabulary development in the sense that we have, the current study substantiates and extends the findings of previous research (Martin & Ellis, 2012; Yang et al., 2017) with active vocabulary learning by high proficient EFL learners. The WM assessed in all of the mentioned studies was executive WM suggesting that despite other research indicating the relation between PTSM and vocabulary learning, the executive WM continues to be a major factor in vocabulary development. In the same way, Linck et al. (2014) found that the correlation between executive WM and L2 processing ( $r = .27$ ) was higher than that of PTSM measures and L2 outcomes ( $r = .17$ ). They suggested that when L2 processing is considered, executive component of WM seems more determinant of L2 learners' performance than PTSM. It might be concluded that CE and PTSM independently contribute to different aspects of vocabulary learning (Martin & Ellis, 2012).

The findings also support the assertion that the high-load tasks lead to more involvement of WM. According to Robinson (2007) and some other studies (e.g. Sagarra & Herschensohn, 2010), with the increase in task load, the WM effect arises, the position which is supported in this study. Therefore, it is not surprising that when passive vocabulary knowledge was concerned, the performance of the high and low WM groups was approximately alike. It is due to the less intrinsic load that was imposed on the attentional resources of the participants who merely picked up the meaning of the annotated words. Regarding active knowledge of the annotated words, the load on the task was higher as the participants had to study the meanings as well as simultaneously learn how they were used in a sentence by the lecturer in the listening text. This task which caused split-attention was better accomplished by the high-WM participants. Correspondingly, considering the findings in the present study suggesting that participants' WMC also had an effect on their comprehension scores, it can be deduced that 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. Subsequently, it can be concluded that executive WM has an influential role in learning vocabulary by advanced learners of English. Consequently, the limited working memory of the learners as well as the load of the information have resulted in low-WM participants' low performance in active vocabulary knowledge.

Another finding was the absence of an interaction between the modality effect and WMC remarked on in other studies (Brunyé, Taylor, Rapp, & Spiro, 2006; Gyselinck et al., 2008; Kozan et al., 2015). The difference between the high and low WM groups was insignificant for passive vocabulary learning regardless of the annotation mode. Comparably, the two WM groups performed differently in the active part of the test with respect to all words either annotated verbally or visually. This disparity might be

due to the fact that while the focus of previous research was on the effect of modality and WM on reading or listening comprehension, we focused on vocabulary knowledge. The lack of interaction between the WMC and the modality effect can also be attributed to the differences in the participants' preferences for verbal or visual illustrations. Some research has demonstrated that as some learners have higher verbal ability and lower visual ability, they perform better when the words are annotated verbally (Kim & Kim, 2012). This preference might have obscured the possible interaction existing between the WMC and the input mode in our study.

Furthermore, knowledge of the text topic has been suggested as an influential factor causing variation in the WMC effect (Alptekin & Erçetin, 2011; Joh & Plakans, 2017; Leaser, 2007). Studies in this domain have found an interaction between WM and topic familiarity. In the present study, the participants' unfamiliarity with the text-topic might have moderated the way they interacted with the listening text. Cowan (2014) stated that inadequate knowledge about the topic of the text accentuates the effect of WM. It holds true in our study, as the participants were unfamiliar with the text topic. The low-WM group were forced to comprehend the text, focus on details given in the text, while acquiring the unknown words in the text which seems to have overloaded their WMC. Provided that the learners' schema about text content had been activated or they had been introduced to some points, the difference in their performance on the active aspect of the VKS tests might have decreased. This notion is verified with the results pointing to the WMC as the predictor of listening comprehension.

### *5.3. Tasks with diverse involvement loads and vocabulary learning*

In an attempt to examine the claims made in the ILH (Laufer & Hulstijn, 2001), we also sought to investigate whether tasks of different involvement loads lead to improved learning. As the overall results partially supported the hypothesis, they are consistent with those of previous studies (Kim, 2011; Min, 2008). The results diverge from the studies which have provided full support for ILH (Feng, 2015; Pourakbari & Biria, 2015). Considering passive vocabulary knowledge, despite the better performance of the Sentence writing group, the Cloze deletion and Paragraph writing groups' performance was roughly the same. This contradicts the basic assumption of the ILH that the increasing amount of involvement load would lead to differential vocabulary learning. That sentence writing task yielded superior performance over the cloze deletion and paragraph writing tasks, despite no differences between the cloze deletion and paragraph writing matches the results of the first settings in Laufer and Hulstijn (2001) and Kim (2011). Also, Martínez-Fernández (2008) found no significant difference between the multiple-choice annotations and control task in initial word learning despite the obvious difference in the load. There are also other research which did not find increased vocabulary development with the increase in evaluation component or with the presence of search (Bao, 2015; Beal, 2007). In line with them, moderate need and evaluation could not augment vocabulary development in the present study. The participants completing the paragraph task (-N, -S, -E) and

those completing the cloze deletion task (+N, -S, +E) performed approximately alike. Strong evaluation, present in the sentence writing task, however, made a big difference in the performance of the groups. Evidently, the nature of the tasks integrates with the components of involvement (Kim, 2011; Yang et al. 2017).

It was observed that the differences between the three tasks (i.e. effect sizes) were higher when the words were visually annotated. This fact implies that an interaction existed between task type and annotation mode. Although this interaction is not great, it indicates that the effect of task type was heightened through dual-mode annotations. Relatedly, it is asserted that including illustrations enhances meaning acquisition; that is passive development (Boers, Warren, He, et al., 2017). Subsequently, the Cloze deletion and Sentence writing groups' word knowledge was consolidated as a result of visual annotations and the tasks they completed. On the same note, the findings sustained that visual aids increase meaning recall in the subsequent tests (Jones & Plass, 2002). This is while Yang et al. (2017) identified significant advantages of both the sentence writing and cloze deletion tasks over the paragraph writing task in improving word knowledge

The results for active vocabulary knowledge are contradictory. When words were textually annotated, the Cloze deletion group outperformed the Paragraph writing group. Yet the difference was rendered insignificant. The sentence writing task, identical to passive part of the VKS, led to the best scores. The opportunity to apply the new words to new sentences is one of justification for this superiority. Nevertheless, visual annotations yielded full support for the ILH. For words under dual-mode annotations, the Cloze deletion group significantly performed better than Paragraph writing group. Apparently, in this aspect of word knowledge, the interaction between the two variables is heightened. One contributing factor might be that the Sentence writing and Cloze deletion groups had the chance to apply the new words in a new context, which strengthened active vocabulary knowledge through dual annotation mode (Schmitt, 2010). Furthermore, the results are in congruence with those studies maintaining that participants completing tasks which require production surpass those who complete other kinds of tasks (Huang, Willson, & Eslami, 2012; Yang et al., 2017). This is due to the fact that such tasks impose greater loads on cognition, which in turn boosts vocabulary development. As the effect of time on task was controlled and the results signified insignificant effect of time on vocabulary development, it is safe to conclude that the extent to which each task contributed to vocabulary gain was solely the result of the nature of the tasks rather than the amount of time spent on completing them.

#### *5.4. Interaction effect of WM, multimedia annotations, and task type on vocabulary learning*

In answering the fourth research question, which explored the interaction effect of WM, annotation mode and task type on vocabulary development, the only significant interaction was on active vocabulary learning through text-only annotations. In the

continuation of previous results, it was observed that the lack of pictorial annotations caused the disappearance of task type effect, namely the cloze deletion and the paragraph writing task effect for the low WM group. As the three post-listening activities led to differences in the performance of the participants on the VKS tests, the results pointed to the significant task type effect for the high WM group. Although no pictorial aid was provided for this set of words in text-only input mode, the high WM group seemed to have grasped not only the meaning but also the usage of the words in a sentence. Therefore, it was the task types that made the difference in the performances of the participants in this group. Conversely, the low WM group appeared to have missed how to use the words appropriately in a sentence, which is probably attributable to their focus on picking up the meanings which were presented in the absence of pictorial aid. This is an interesting finding as previous research has found a split attention phenomenon when visual and written data are presented simultaneously (Seufert, Schütze, & Brünken, 2009). In the current study, on the other hand, it was observed that the concurrent presentation of an auditory text and the written meanings of the new words restricted the low WM group's attention to the meaning, which momentarily seemed more important, rather than how it was employed in a sentence by the lecturer. Moreover, the provision of meanings without a picture, which decelerated the grasp of the meaning, together with no opportunity to further practice the new words in a new context in the cloze deletion and the paragraph writing tasks resulted in poor performance of these two groups in the production part of the test. Consequently, we can, albeit conservatively, extend the split attention effect to auditory plus visual presentations.

The reason why this interaction was not observed for active vocabulary learning for words annotated through both text and picture is most likely a result of the speed of analysis the pictures provided the participants with. With the text-only modes, the low WM participants had to focus on the meaning for as long as it was on the screen to learn it, thus they missed some parts of the listening text as the load on their cognitive system exceeded their available resources (Sweller, 2010). This is while the text-plus-picture annotations accelerated and deepened the learning process and the pictures representing the meanings offloaded the cognitive system (Mayer, 2014). It seems that when the words were learnt through multimedia presentation, (i.e. passive knowledge through both input modes and active knowledge under dual-mode annotations), the task type effect was heightened, but when the words were not fully acquired at first, this effect diminished.

## 6. Conclusions

The present study shows how individual differences; that is, WMC and the modality effect plus tasks with differences in their involvement loads affect passive and active development of proficient EFL learners. The findings implied that the modality effect may hold true for both passive and active vocabulary learning in a system-paced instruction and WM signifies at least one portion of the differences in the performance. Surprisingly, WMC did not, however, mediate the modality effect on

vocabulary learning. The overall results implied the positive effect of word-focused post-listening activities. The ILH was partially supported, as the sentence writing task with the highest involvement load produced the best performance in both passive and active knowledge. The cloze deletion and the paragraph writing tasks, despite different involvement loads, did not bring about different performances in the VKS tests. Accordingly, the results verify that the “evaluation” component of the ILH is the most significant factor in vocabulary learning (Kim, 2011).

Our study also revealed that except for one dimension of vocabulary knowledge (active knowledge under single-mode annotations), no interaction among the three variables was evident. Nevertheless, due to the lack of research on the interaction of these three variables, the need for further research on the subject is imperative. It is apparent that by presenting the materials through dual-mode annotations plus providing post-listening activities, we would enhance the passive and active knowledge of the learners. The post-listening activities which require production seem to benefit learners better, regardless of WM capacities, as these activities engage learners in absorbing different aspects of the new words (Yang et al. 2017).

To sum, the present study aimed to fill the gaps in the literature on the relationships between the executive WM, the modality effect, task-induced involvement load, and vocabulary learning. It constitutes a significant complement to the previous research in that it extended the effect of CE to passive and active vocabulary learning through listening and it also brought forward the effect of post-listening word-focused activities and their interaction with the modality effect and CE on vocabulary acquisition of proficient EFL learners. The pedagogical implication of the current study is that EFL teachers and material designers should bear in mind individual differences, WMC in particular, when creating or choosing materials for different courses. Besides, the materials are best presented with accompanying visual aids, and the activities implemented to enhance vocabulary learning should involve tasks that induce higher loads and require production of the new vocabulary items in new contexts.

### Acknowledgements

This paper is based on corresponding author’s dissertation (Salva Kazemipour Khabbazi, 2021).

### 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. <https://dx.doi.org/10.1080/09588220802090246>
- Alptekin, C., & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly*, 45, 235-266. <https://dx.doi.org/10.5054/tq.2011.247705>

- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
- 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.
- Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System*, 53, 84–95. <http://dx.doi.org/10.1016/j.system.2015.07.006>
- Beal, V. (2007). The weight of involvement load in college level reading and vocabulary tasks. Unpublished doctoral dissertation, Concordia University.
- Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning*, 30(7), 709–725. <https://doi.org/10.1080/09588221.2017.1356335>
- Boers, F., Warren, P., He, L., & Deconinck, J. (2017). Does adding pictures to glosses enhance vocabulary uptake from reading? *System*, 66, 113–129. <https://dx.doi.org/10.1080/09588221.2017.1356335>
- Brunfaut, T., & Révész, A. (2014). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, 49(1), 141–168. <https://doi.org/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. <https://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. <https://doi.org/10.1017/S0958344017000155>
- 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://doi.org/10.3758/BF03196772>
- Cottam, M. E. (2011). The effects of visual and textual annotations on Spanish listening comprehension, vocabulary acquisition and cognitive load. Unpublished doctoral dissertation, Arizona State University. Retrieved from <https://www.learntechlib.org/p/118131>
- Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, 26, 197–223. <https://dx.doi.org/10.1007/s10648-013-9246-y>
- 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.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://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>.
- De Westelinck, K., Valcke, M., De Craene, B., & Kirschner, P. (2005). Multimedia learning in social sciences: Limitations of external graphical representations. *Computers in Human Behavior*, 21, 555–573. <https://doi.org/10.1016/j.chb.2004.10.030>
- Doolittle, P., Terry, K., & Mariano, G. (2009). Multimedia learning and working memory. In R. Zheng (Ed.), *Cognitive effects of multimedia learning* (pp. 17–33). New York: Information Science References.
- Feng, T. (2015). Involvement load in translation tasks and EFL vocabulary learning. *The New English Teacher*, 9(1), 83–101.

- File, A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly*, 44, 222–249. <https://doi:10.5054/tq.2010.219943>
- French, L. & O'Brien, I. (2008) Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29, 463–487. <https://doi:10.1017/S0142716408080211>
- Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology*, 22, 353–374. <https://doi.org/10.1002/acp.1411>
- Huang, S., Willson, V., & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal*, 96, 544–557. <https://doi.org/10.1111/j.1540-4781.2012.01394.x>
- Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). The median split: Robust, refined, and revived. *Journal of Consumer Psychology*, 25(4), 690–704. <https://dx.doi.org/10.1016/j.jcps.2015.06.014>
- 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. <https://doi:10.1111/bjet.12129>
- Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System*, 70, 107–120. <https://dx.doi.org/10.1016/j.system.2017.07.007>
- 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. <https://doi.org/10.1111/1540-4781.00160>
- Juffs, A., & Harrington, M. W. (2011). Aspects of working memory in L2 Learning. *Language Teaching: Reviews and Studies*, 42, 137–166. <https://doi:10.1017/S0261444810000509>
- 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. <https://doi:10.1037/0096-3445.133.2.189>
- Keating, G. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, 12(3), 365–386. <https://doi:10.1177/1362168808089922>
- Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 61, 100–140. <https://doi.org/10.1111/j.1467-9922.2011.00644.x>
- Kim, D., & Kim, D. (2012). Effect of screen size on multimedia vocabulary learning. *British Journal of Educational Psychology*, 43(1), 62-70. <https://doi:10.1111/j.1467-8535.2010.01145.x>
- 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. <https://dx.doi.org/10.1016/j.system.2015.09.001>
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review*, 59(4), 567–587. <https://doi:10.3138/cmlr.59.4.567>
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5, 223–250. <https://doi.org/10.1075/eurosla.5.11lau>
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26. <https://doi:10.1093/applin/22.1.1>

- 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. <https://dx.doi.org/10.1111/j.1467-9922.2007.00408.x>
- Lin, C., & Tseng, Y. (2012). Videos and animations for vocabulary learning: A study on difficult words. *The Turkish Online Journal of Educational Technology*, 11(4), 346–355.
- Linck, J.A., Osthus, P., Koeth, J.T. & Bunting, M.F. (2014). Working memory and second language comprehension and production: a meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861–883. <https://doi:10.3758/s13423-013-0565-2>
- 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. <https://doi:10.1111/j.1467-9922.2010.00565.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. <https://doi:10.1017/S0272263112000125>
- Martínez-Fernández, A. (2008). Revisiting the involvement load Hypothesis: Awareness, type of task and type of item. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected proceedings of the 2007 second language research forum* (pp. 210–228). Somerville, MA: Cascadia Proceedings Project.
- 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. [https://doi:10.1207/S15326985EP3801\\_6](https://doi:10.1207/S15326985EP3801_6)
- Min, H. T. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, 58(1), 73–115. <https://doi:10.1111/j.1467-9922.2007.00435.x>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Harlow: Longman.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.
- 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.
- Pichette, F., de Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, 33(1), 66–82. <https://doi:10.1093/applin/amr037>
- Phillips, D. (2015). *Longman preparation course for the TOEFL iBT test*. White Plains, NY: Pearson Education.
- Pourakbari, A. A., & Biria, R. (2015). Efficacy of task-induced involvement in incidental lexical development of Iranian senior EFL students. *English Language Teaching*, 8(5), 122–131. <http://dx.doi.org/10.5539/elt.v8n5p122>
- Rassaei, E. (2017). Computer-mediated textual and audio glosses, perceptual style and L2 vocabulary learning. *Language Teaching Research*, 1–19. Retrieved online on May, 2017 from <https://journals.sagepub.com/doi/abs/10.1177/1362168817690183>
- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL*, 45, 193–213. <https://doi:10.1515/iral.2007.009>



- Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120, 2022–2039. <https://dx.doi.org/10.1016/j.lingua.2010.02.004>
- Sanchez, C. A. & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34(2), 344–355. <https://doi.org/10.3758/BF03193412>
- Schmitt, N. (2008). Review article: instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363. <https://doi:10.1177/1362168808089921>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.
- Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude-treatment-interaction study. *Learning & Instruction*, 19, 28–42. <https://doi:10.1016/j.learninstruc.2008.01.002>
- 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>
- Tabbers, H. K., & de Koeijer, B. (2010). Learner control in animated multimedia instructions. *Instructional Science*, 38(5), 441–453. <https://doi:10.1007/s11251-009-9119-4>
- Williams, J. N. (2011). Working memory and SLA. In S. M. Gass, & A. Mackey (Eds.), *The handbook of second language acquisition* (pp. 427–441). New York, NY: Routledge.
- 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. <https://doi:10.1016/j.system.2017.09.012>
- Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, 24(1), 39–58. <https://dx.doi.org/10.1080/09588221.2010.523285>

## Appendix A. Text-plus-picture Annotations





### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).