



Task Types Effects and Task Involvement Load on Vocabulary Learning of EFL Learners

Parviz Alavinia

Dr., corresponding author, English Language Department, Urmia University, Urmia, Iran, pevinia2006@yahoo.com

Hanieh Rahimi

PhD Candidate in TEFL, English Language Department, Urmia University, Urmia, Iran, ha.rahimi1983@gmail.com

This study investigated the impact of task types on English learners' receptive and productive vocabulary knowledge. A total of 125 (70 female and 55 male) first year ELT students from eight intact classes taking Academic Reading Course in the winter semester of 2017-2018 were assigned to four experimental groups randomly and a control group randomly assigned to one of five tasks of learning 45 target words. The design of the tasks was based on the involvement load hypothesis (ILH) arguing that learning of unfamiliar words to be contingent on the amount of task induced involvement. The components of involvement in ILH include need (N), search (S), and evaluation (E). In this study, the tasks induced the same or different involvement loads regarding the presence and strength of each component: writing (+N, +S, ++E), combining (+N, -S, +E), fill in the blank (+N, -S, +E), definition (+N, -S, +E), and control (-N, -S, -E). After the last treatment session, both receptive and productive knowledge of the target words were measured. Moreover, an unexpected delayed post-test was administered 1 month later. The results revealed that all output tasks were more effective than the control task in enhancing the participants' receptive and productive vocabulary knowledge.

Keywords: involvement load hypothesis, output task, receptive vocabulary knowledge, productive vocabulary knowledge, vocabulary learning, EFL learners

INTRODUCTION

Vocabulary learning has been a challenging task for Iranian EFL learners for decades. Forgetting and inability to use the learnt words productively are their common problems. In order to assist Iranian English teachers and material developers to satisfy the learners' needs, this study put the Involvement Load Hypothesis into practice. It emphasizes the

Citation: Alavinia, P. & Rahimi, H. (2019). Task Types Effects and Task Involvement Load on Vocabulary Learning of EFL Learners. *International Journal of Instruction*, 12(1), 1501-1516. <https://doi.org/10.29333/iji.2019.12196a>

impact of a task's determining factor for improving vocabulary learning including a combination of need, search, and evaluation of the new words, which is called task involvement load.

This study emphasizes the impact of a task's determining factor for improving vocabulary learning including a combination of need, search, and evaluation of the new words, which is called task involvement load. It is necessary for language researchers to explore why certain tasks are more effective than others in L2 vocabulary acquisition. One claim has been that the acquisition of new words is related to the degree of involvement with a task. This is known as the Involvement Load Hypothesis (ILH, Laufer & Hulstijn, 2001). The ILH predicts that the greater the demands that the word learning task places on the second language learner, i.e. the greater its involvement load, the more likely the word will be learned. Because of the importance of vocabulary learning and retention, Hulstijn and Laufer proposed the notion of the Involvement Load Hypothesis in 2001. This hypothesis consists of three main components: Need, Search, and Evaluation. This hypothesis claims that tasks inducing higher involvement load produce better vocabulary retention effects. Accordingly, this study intended to find out whether word learning and retention in a second language is contingent upon a task's Involvement Load Hypothesis, i.e. the amount of need, search, and evaluation.

REVIEW OF THE LITERATURE

Vocabulary is widely regarded as a crucial part of language learning by virtually all second language (L2) learners and their teachers. For both language teachers and learners vocabulary is obviously a top priority (Nation & Wang, 1999; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb, 2007, Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Peters, 2007; Pichette, De Serres, & Lafontaine, 2012).

The ILH, as a major development in field of L2 vocabulary research, has received a great deal of attention as it is clear, precise, and can be operationalized (Bao, 2015; Keating, 2008; Tang & Treffers- Daller, 2016; Zou, 2017). The ILH consists of three components: need, search, and evaluation. Need is a noncognitive but motivational factor. Search is the relationship between form and meaning of unknown words. Evaluation includes making a decision on the appropriate word with its related meaning in context. These elements are authorized regarding their distinction. If a component is absent (–) the score is 0. If the component is moderate (+), the score is 1 and the component gains 2 if the involvement is strong. It is believed that the best result in learning new vocabulary is obtained through a task with the highest degree of involvement load (Marmol & Sanchez-Lafunte, 2013).

The ILH argues that tasks with three constructs of need (N), search (S), and evaluation (E) have more effectiveness on vocabulary learning than tasks with lower involvement loads (Laufer & Hulstijn, 2001). Need is a motivational construct dealing with the “need to achieve” (Laufer & Hulstijn, 2001, p. 14), while search and evaluation are cognitive constructs having to do with paying attention to form–meaning relationship. Tasks may induce these involvement elements to three possible degrees: none, moderate, and strong. A task involvement load is referred to as the combination of these involvement

elements, which can be absent or present, moderate or strong (Bao, 2015). A moderate involvement is given an index of 1 and a strong involvement receives an index of 2. It is argued that “the higher the scores of need, search, and evaluation are, the greater the involvement load in learning an unknown word is” (Bao, 2015, p. 85).

In testing the ILH, much research was done to compare several word learning tasks assumed to have different involvement loads (e.g., Bruton, 2007; Huang & Lin, 2014; Hulstijn & Laufer, 2001). The ILH also suggests that, regardless of word learning task type, the same presence of an involvement component always leads to the same amount of word learning. Nevertheless, little research was done to investigate this issue, either (i.e., Folse, 2006; Kim, 2008). For these reasons, this study investigated whether several word-focused output tasks were all more effective in English as a Foreign Language (EFL) learners' receptive and productive vocabulary acquisition than a control task (with no focus on the target words). As vocabulary knowledge plays a crucial role in L2 learning, it is necessary for teachers to select and use proper vocabulary learning tasks (Bao, 2015).

Martínez-Fernández (2008) measured the impact of three tasks on vocabulary development. Vidal (2011) also compared the effects of the listening and reading tasks on vocabulary acquisition across levels of proficiency. Participants either read the texts or watched the lectures for learning unknown words. The result indicated the superiority of the reading task over the listening task for vocabulary acquisition.

Vidal (2011) also compared the effects of the listening and reading tasks on vocabulary acquisition across levels of proficiency. Participants either read the texts or watched the lectures for learning unknown words. The result indicated the superiority of the reading task over the listening task for vocabulary acquisition. However, the difference between the impacts of the tasks decreased as the participants' proficiency increased. The reading task's superiority was also evident in the delayed vocabulary test.

Kim (2008) also provided empirical evidence for the involvement load hypothesis in a carefully designed study consisting of two experiments. The first experiment addressed the effectiveness of three vocabulary tasks with different levels of involvement index. The second experiment, on the other hand, examined whether tasks with equal involvement load would lead to equivalent initial and later retention of words by 20 adult ESL learners at two different levels of proficiency. In line with other studies, the results showed that a higher involvement index leads to more effective initial and delayed vocabulary learning. Furthermore, Kim found that identical involvement index in two tasks unfolded similar results for the two L2 proficiencies. Despite the overall support, Folse (2006) reports that his study showed word learning to be more a function of repeated exposure than involvement.

The significant effect of increasing the presence of evaluation was proven in other studies as well. The results suggested that tasks with higher ILLs were found to be better in promoting incidental vocabulary acquisition (Behbahani, Pourdana, Maleki & Javanbakht, 2011; Feng, 2015; Sarbazi, 2014).

In addition to studies testing the adjustment of evaluation component only, there have been studies testing the effect of search and evaluation components as well. Sarani, Negari and Ghaviniat (2013) designed six tasks with varying involvement loads. Three of the tasks were receptive and the other three were productive. Such a design would help investigating whether different tasks (receptive, productive) with similar involvement load levels would yield similar results in terms of incidental VG and VR. Receptive tasks were true-false (moderate need), matching (moderate need and moderate evaluation) and multiple choice (moderate need, search and moderate evaluation) inducing involvement load indexes of 1, 2, and 3 respectively. Using a similar design to that Sarani et al. (2013) used; Ghabanchi, Davoudi and Eskandari (2012), Hazrat (2015) and Pourakbari and Biria (2015) also found similar results proving the significant effect of ILL on vocabulary acquisition. On the other hand, there were also contradictory results which suggested that increasing the level of evaluation (Bao, 2015; Beal, 2007; Jahangiri & Abilipour, 2014; Keyvanfar & Badraghi, 2011), and inserting search (Haratmeh, 2012; Jahangard, 2014; Marmol & Sanchez-Lafuente, 2013) didn't always result in increased vocabulary gain. Participant related factors such as attention span, writing skills and dictionary use skills were found to be hindering the effect of ILL.

Rather than using passage-context designs, a few studies used sentence-context designs to test the ILH (e.g. Folsie, 2006; Pichette, de Serres, & Lafontaine, 2012; Webb & Kagimoto, 2009). Very few studies (e.g., Webb, 2005, in the second experiment) fully supported the ILH. Folsie (2006) suggested that the important feature of a given L2 vocabulary exercise may not be involvement load but the number of word retrievals required. Webb (2005) found that confirming or disconfirming the ILH also depends partly on how tasks are designed, how task time is handled and how vocabulary knowledge is measured.

Similarly, Pichette et al. (2012) indicated that, when the ILH is tested, word factors like concreteness need to be taken into account. Webb and Kagimoto (2009) even argued that task difficulty and learning strategies may contribute to the different results for the higher and lower level L2 learners.

Keating (2008), for example, investigated the effects of three tasks with different involvement loads on vocabulary retention of 79 Spanish learners. To this end, three tasks with different involvement loads were selected: reading comprehension task (involvement load of 1), reading plus fill-in task (involvement load of 2), and sentence writing task (involvement load of 3). The results showed that the participants could learn more words in sentence writing task with the highest involvement load. Nevertheless, Keating (2008) did not include any control group in his study.

In another study, Marmol and Sanchez-Lafuente (2013) studied the effects of four types of tasks on EFL vocabulary learning. The participants were 28 primary school English as a second language (ESL) learners in Spain. Eighteen words including six nouns, six adjectives, and six verbs, were selected randomly from a short story. The participants were assigned to four different tasks with different involvement loads: reading comprehension with marginal glosses, reading comprehension and gap-filling, writing

with marginal glosses, and writing and dictionary use. All the participants took a receptive and a productive vocabulary test. The results showed that doing a task with the highest involvement led to the best result in L2 vocabulary learning. Their study, however, lacked a control group, and the sample size was small.

Some researchers (e.g., Folse, 2006; Webb, 2005; Zou, 2017) have laid emphasis on writing tasks as effective ones in vocabulary learning. In the same vein, Feng (2014) included sentence writing task in his study. Feng (2014) examined the effects of three translation tasks on EFL learners' vocabulary learning based on the ILH. In this study, 30 verbs were selected from business documents to be taught to 60 EFL learners via three different translation tasks: translation- only mode, translation plus fill-in exercises, and translation plus sentence writing. The result indicated sentence writing could significantly improve passive and active word learning and retention, whereas translation-only task had the lowest effect. However, Feng's (2014) study lacked a control group, did not control the word type, and sufficient information for power analyses was not provided.

Tang and Treffers-Daller (2016) have recently examined the effects of different tasks on L2 incidental vocabulary learning based on the predictions of ILH. To this end, 230 Chinese EFL learners whose proficiency was at A2 level on the Common European Framework of Reference for languages (CEFR) were selected. Six different tasks with different involvement loads were designed. The results showed that tasks with a higher involvement load were significantly better than tasks with a lower involvement load both in the immediate and the delayed posttests. However, like most previous studies, this study suffers from lack of a control group and insufficient information for power analyses.

In Iranian context, few studies have been conducted on the effect of task types and involvement indices on EFL learners' vocabulary learning. Yaqubi, Rayati, and Allemzade Gorgi (2010), for instance, randomly assigned 60 EFL learners to three groups: Group 1 completed an input-oriented task with an involvement load of 3, Group 2 was given the same type of task but with an involvement load of 2, and Group 3 completed an output-oriented task with the involvement load of 3. The results were contrary to the prediction of the ILH, that is, Task 2 was superior to Task 1, which had a higher index. Moreover, the learners who had completed Task 3 significantly did better than those who did Task 1, despite their index equivalency. However, this study lacked a control group, had a small sample size, provided no information for power analyses, and did not measure the participants' productive knowledge.

RESEARCH QUESTIONS AND HYPOTHESES

The present study was an effort to investigate the following research questions and hypotheses:

RQ1: Are different output tasks (i.e., writing, definition, combining, and fill in the blank) all conducive to EFL receptive vocabulary knowledge?

RQ2: Are different output tasks (i.e., writing, definition, combining, and fill in the blank) all conducive to EFL productive vocabulary knowledge?

RQ3: Which output task (i.e., writing, definition, combining, and fill in the blank) will be more effective in EFL receptive vocabulary knowledge?

RQ4: Which output task (i.e., writing, definition, combining, and fill in the blank) will be more effective in EFL productive vocabulary knowledge?

H01: Tasks with different involvement loads do not lead to differential gains in immediate receptive vocabulary knowledge.

H02: Tasks with different involvement loads do not lead to differential gains in delayed receptive vocabulary knowledge.

H03: Tasks with different involvement loads do not lead to differential gains in immediate productive vocabulary knowledge.

H04: Tasks with different involvement loads do not lead to differential gains in delayed productive vocabulary knowledge.

METHOD

Participants of the Study

The participants were first-year non-English majors from five intact English classes at Azad University, and each class was randomly assigned to one of the four vocabulary learning tasks. First, a total of 165 (94 female and 71 male) freshman students in Azad University participated in this study. They were selected from three departments in Azad University, in Tehran, Iran. The participants' age ranged between 18 and 25 the participants were equally assigned to five vocabulary learning tasks ($n = 25$ each) through simple random sampling. However, after the delayed post-test, the number of the participants was reduced to 125 (70 female and 55 male). Finally, the number of participants at each vocabulary learning task was as follows: writing ($n = 25$), definition ($n = 25$), fill in the blank ($n = 25$), combining ($n = 25$), and control ($n = 25$).

As Oxford Placement Test (OPT) is easy to administer and has a well-established reliability and validity (Allan, 2004), it was adopted in this study. OPT includes two sections, grammar and listening, each of which consists of 100 items. The required time to complete the test is 60 min. Each correct item received 1 point. Therefore, the maximum possible score was 200. The participants' scores ranged from 105 to 118, suggesting that they were at the elementary level of English proficiency.

Instructional Materials

Target Words. In this study, first, 45 English target words, with equal parts of speech, were randomly selected from the *Inside Reading 1* (Burgmeier, Arline, 2013). This book includes 10 Units and 5 words were randomly selected from each lesson. The reasons behind selecting equal parts of speech were controlling the possible confusion of part of speech with task type and enhancing the generalizability of the findings. Moreover, the target words were selected from *Inside Reading* whose words are a little beyond the participants' proficiency level.

To select the target words, a test was also administered to 60 nonparticipants at another university. The testees were asked to write the Persian equivalents of the words. The nonparticipants were also at the elementary level as the participants in the experiment.

After administering the test, 30 words (i.e., 10 nouns, 10 verbs, and 10 adjectives) unknown to all the testees were selected as the target words (see Appendix A).

EFL reading task

In this study, 30 sentences were adapted from Inside Reading 1 (Burgmeier, Arline, 2013) in which each target word was embedded (see Appendix B). The teacher-researchers and two other EFL teachers checked the appropriacy of vocabulary and syntax of the reading sentences for the participants. The 45 sentences were randomly divided into three sets; each set was presented in one session.

Each set included 10 sentences and 10 target words. After each sentence, the corresponding gloss for the target word was provided in the brackets. All the groups received the sentences in the same random order.

Vocabulary knowledge test

This study adopted Min's (2008) four-item Vocabulary Knowledge Scale (VKS) in which the unknown and known word categories are separated (see Appendix C). "Categories of VKS offer no clues to the target words and thus can more accurately reflect the students' knowledge about the target words" (Min, 2008, p. 85). In this study, both receptive and productive vocabulary knowledge of participants were measured. As a result, Categories III and IV in the VKS measured the EFL receptive and productive vocabulary knowledge, respectively.

EFL vocabulary learning tasks

There were three sets of vocabulary learning exercises, and each set was presented, immediately following each set of sentences in the reading task wherein the test words appeared. All the definitions/descriptions of the test words for the control and definition groups were chosen or adapted from the foregoing dictionaries. The difficulty level of each selected or adapted sentence was deemed appropriate for the participants in terms of vocabulary and syntax. The test words were put in boldface, meaning that they appeared in the reading task and could be consulted.

The control task consisted of meaning matching exercises, whose test words were not the target ones. For each set of matching exercises, six test words (chosen from the preceding reading task) were presented on the left, and eight definitions/ descriptions were given on the right. The control group needed to draw a line connecting each test word to a proper definition/ description. As a basic design principle, the words in each definition/description were generally more frequently used than the test words. The test format where the definitions/descriptions outnumbered the test words was meant to minimize the participants' reliance on guessing.

Procedure

This study followed the procedure of Bao's (2015) study. This study adopted a sentence-context between-subjects design to investigate how task type affected EFL vocabulary knowledge. The study designed four tasks, i.e., control, definition, combining, filling the blank and writing. Task type was the between-subjects factor. EFL vocabulary knowledge was the dependent variable and measured in terms of

receptive and productive vocabulary knowledge. First, the participants were randomly assigned to four experimental groups receiving a vocabulary learning task (i.e., writing, definition, fill in the blank, and combining) and a control task. The same or different involvement loads were utilized for all tasks except for the control task to test the contribution of each task type to vocabulary learning. The writing task induced involvement load index of 4 (+N, +S, ++E), and combining, fill in the blank, and translation tasks all induced an involvement load index of 2 (+N, -S, +E).

During the experiment, each class of participants needed to finish the sentence reading task and one of the five vocabulary learning tasks within the maximum time limit as required in the instructions. All the participants were allowed to consult the reading task for the glosses of the target words while performing the vocabulary learning tasks. The writing group was required not to copy the sentences in the reading task when making sentences. The purpose of this requirement was to force them to pay attention to the meaning, part of speech and usage of each target word. They were allowed to hand in the materials ahead of the scheduled time.

After the teachers collected all the materials, there was a ten minutes' break. Then, without announcing in advance that the participants would take vocabulary knowledge post-test, the teachers administered the test. After the 25-min test was over, the teachers collected all the test sheets.

Two weeks before the study, all participants took the VKS. In the main study, the experimental groups were taught the target words by the teacher-researchers in three sessions. In each session, 10 target words were taught through a vocabulary learning task presented immediately following a reading task in which the target words appeared. For the control group, the teacher-researchers gave the task which included meaning matching exercises in which the test words were not the target words. For each set of exercise, 10 test words had to be matched with 12 definitions. The number of the definitions was more than the test words to decrease the participants' guessing.

For the writing tasks, the participants were required to write semantically acceptable and grammatically correct sentences or in 15 min. The teacher-researchers were available to answer any questions and the participants had access to dictionaries. In both groups, the participants did the tasks individually. For the combining task, each sentence was segmented into separate parts. The combining group was required to combine all segments into a grammatically correct sentence.

For the fill-in the- blank task, the participants were to fill in the blanks with appropriate given (target) words. At the end of the last treatment session, the immediate post-test was administered in 20 min. To assess the participants' long-term retention of the target words, an unexpected delayed post-test was also administered 1 month after the experiment.

For the other tasks, three sets of vocabulary learning exercises were presented in the same order as for the control task, but the test requirements were different. For the definition task, each set of six definitions/descriptions was provided, and each corresponding target word should be written down in the preceding blank. For the combining task, each sentence in the three sets was segmented into five word

combinations. The target words and their collocates (if any) were put together as one segment. The combining group needed to combine all segments in a test item into a grammatically correct sentence.

Scoring and data analysis

To answer the research questions the following statistical analyses were used. To compute all data, first the mean and standard deviation of each individual test was calculated. Then, the performances of the groups in all tests were compared to check whether there were significant differences among different types of tasks. Finally, it was compared with its corresponding test of the other group to find which type of task is more beneficial in learning and retention of the target words.

Two experienced EFL teachers were trained to score the participants' receptive and productive knowledge of each target word. They independently scored each participant's responses for Categories III and IV. The inter-rater agreement was 100%, as all disagreements were discussed between the two raters until consensus was reached.

In this study, Min's (2008) scoring was followed to measure the participants' receptive and productive vocabulary knowledge of the target words. For Categories III and IV, 1 point or 0 point was given. Category III measured the receptive knowledge of a target word. If the given synonym was improper, or no response was given, 0 point was considered. A correct synonym also received 1 point. Category IV measured the productive knowledge of a target word. If no point was given to Category III, 0 point was given to Category IV. If a target word's meaning was inappropriate or ungrammatical in the sentence context, 0 point was given. One point was given when a target word was both semantically and grammatically correct in the sentence albeit other parts of the sentence had errors.

The participants' responses were scored by two experienced EFL teachers. Each participant's responses were independently scored. In the pre-test, the Cronbach Alpha reliability indices for the receptive and productive tests were 0.97 and 0.94, respectively. Then, the two raters discussed all discrepancies in scoring until they reached a unanimous agreement. Hence, the inter-rater agreement was 100% in the immediate and delayed post-tests. To address the research questions, a mixed 3 . 6 ANOVA, one-way ANOVA, and Tukey post hoc tests were run on receptive and productive test scores separately. Significance level was set at .05.

FINDINGS

First, normality of data was measured through Komologroph Smirov test. All significant values in the test were above the significance level of .05, implying normal distribution of data. The results also revealed that the writing group had the highest mean scores on both receptive and productive measures of vocabulary knowledge in immediate and delayed post-tests.

Then, a mixed 3 . 6 ANOVA, with time (i.e., pre-test, immediate post-test, and delayed post-test) and task type (i.e., writing, definition, fill in the blank, combining, and control) as two main factors, was done. A one-way ANOVA along with post hoc Tukey

tests was also conducted for the overall comparison of the five groups on all tests. Results of mixed ANOVA on the receptive vocabulary tests are displayed in Table 1.

Table 1
One-Way ANOVA Delayed Post-Test & Immediate Post-Test Receptive

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	663.344	9	73.705	22.923	.000
Within Groups	771.680	240	3.215		
Total	1435.024	249			

Table 2
One-Way ANOVA Delayed Post-Test & Pre-Test Receptive

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	868.564	9	96.507	22.226	.000
Within Groups	1042.080	240	4.342		
Total	1910.644	249			

Table 3
One-Way ANOVA Immediate Post-Test & Pre-Test Receptive

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	727.156	9	80.795	16.078	.000
Within Groups	1206.080	240	5.025		
Total	1933.236	249			

The results show significant main effects of task type ($F_{2, 125} = 16.078, p = .000$).

To further examine the differences between the groups, a one-way ANOVA was performed for each receptive test. Significant difference was found in the pre-test ($F, 125 = 16.078, p = 0.000$). In the immediate post-test ($F, 125 = 22.923, p = .000$) and the delayed post-test ($F, 125 = 22.226, p = .000$), significant differences were found. Table 4 illustrates the findings of post hoc analyses on the post-tests.

Table 4
Results of Tukey Post Hoc Tests on Receptive Post-tests

(I) Tasks	(J) Tasks	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Writing Immediate Post-test Receptive	Definiton Immediate Post-test	3.000*	.634	.000	.97	5.03
	Fill Blanks Immediate Post-test	2.920*	.634	.000	.89	4.95
	Combining Delayed Post-test	3.400*	.634	.000	1.37	5.43
	Control Immediate Post-test	5.000*	.634	.000	2.97	7.03
	Writing Pre-test Receptive	5.960*	.634	.000	3.93	7.99
	Definition Pre-test Receptive	5.760*	.634	.000	3.73	7.79
	Fill Blanks Pre-test Receptive	4.840*	.634	.000	2.81	6.87
	Combining Pre-test Receptive	4.800*	.634	.000	2.77	6.83
	Control Pre-test Receptive	5.160*	.634	.000	3.13	7.19

(I) Tasks		(J) Tasks		Mean Difference (I-J)	Std. Error	Sig.	Tukey HSD 95% Confidence Interval	
							Lower Bound	Upper Bound
Writing Delayed Post-test Receptive	Definiton	Delayed	Post-test	2.480*	.507	.000	.86	4.10
	Fill	Blanks	Delayed	2.400*	.507	.000	.78	4.02
	Combining	Delayed	Post-test	2.800*	.507	.000	1.18	4.42
	Control	Delayed	Post-test	5.080*	.507	.000	3.46	6.70
	Writing	Immediate	Post-test	.120	.507	1.00	-1.50	1.74
	Definition	Immediate	Post-test	3.120*	.507	.000	1.50	4.74
	Fill	Blanks	Immediate	3.040*	.507	.000	1.42	4.66
	Combining	Immediate	Post-test	3.520*	.507	.000	1.90	5.14
	Control	Immediate	Post-test	5.120*	.507	.000	3.50	6.74

Table 4 shows that on the immediate post-test, the writing group did significantly better than all groups. All groups also significantly outperformed the control group. Post hoc tests also showed that writing group obtained significantly higher scores than the combining and control groups on the delayed post-test ($p = .000$). Moreover, all groups significantly outperformed the control on the delayed post-test ($p = 000$) except the writing group in immediate post-test and delayed post-test.

A mixed 3.6 ANOVA (i.e., pre-test, immediate post-test, and delayed post-test) and task type (i.e., writing, definition, fill in the blank, combining, and control) as two main factors, was also conducted on the productive vocabulary tests. In addition, a one-way ANOVA along with post hoc Turkey tests were run for the overall comparison of the five groups on all tests. Results of mixed ANOVA on the productive tests are shown in Table5.

Table 5

One-Way ANOVA Delayed Post-Test & Immediate Post-Test Productive

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	836.884	9	92.987	37.012	.000
Within Groups	602.960	240	2.512		
Total	1439.844	249			

Table 6
One-Way ANOVA Pre-test & Delayed Post-Test Productive

Pre-Test Scores					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	779.124	9	86.569	24.076	.000
Within Groups	862.960	240	3.596		
Total	1642.084	249			

Table 7
One-Way ANOVA Pre-test & Immediate Post-Test Productive

Pre-Test Scores					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	657.056	9	73.006	19.195	.000
Within Groups	912.800	240	3.803		
Total	1569.856	249			

The results showed significant main effects of task type ($F, 125 = 19.195, p = .000$). To further examine the differences between the groups, a one-way ANOVA was performed for each receptive test. In the immediate post-test ($F, 125 = 37.012, p = .000$) and the delayed post-test ($F, 125 = 24.076, p = .000$) significant differences were found. Table 8 depicts the findings of post hoc analyses on the post-tests.

Table 8
Results of Tukey Post Hoc Tests on Productive Post-tests

Pre-Test Scores Tukey HSD						95% Confidence Interval	
(I) Task	(J) Tasks	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound	
Writing Immediate Post-test productive	Writing Pre-test productive	6.080 [*]	.552	.000	4.32	7.84	
	Definiton Pre-test Productive	5.040 [*]	.552	.000	3.28	6.80	
	Fill Blanks Pre-test productive	4.600 [*]	.552	.000	2.84	6.36	
	Combining Pre-test productive	4.560 [*]	.552	.000	2.80	6.32	
	Control Pre-test productive	5.040 [*]	.552	.000	3.28	6.80	
	Definition Immediate Post-test productive	3.280 [*]	.552	.000	1.52	5.04	
	Fill Blanks Immediate Post-test productive	3.240 [*]	.552	.000	1.48	5.00	
	Combining Immediate Post-test productive	3.760 [*]	.552	.000	2.00	5.52	
	Control Immediate Post-test productive	5.440 [*]	.552	.000	3.68	7.20	
Pre-Test Scores Tukey HSD						95% Confidence Interval	
(I) Tasks Pre-Test	(J) Tasks Pre-Test	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound	
Writing Delayed Post-test productive	Writing Pre-test productive	6.480 [*]	.536	.000	4.77	8.19	
	Definiton Pre-test Productive	5.440 [*]	.536	.000	3.73	7.15	
	Fill Blanks Pre-test productive	5.000 [*]	.536	.000	3.29	6.71	
	Combining Pre-test productive	4.960 [*]	.536	.000	3.25	6.67	
	Control Pre-test productive	5.440 [*]	.536	.000	3.73	7.15	
	Definition Delayed Post-test productive	3.120 [*]	.536	.000	1.41	4.83	
	Fill Blanks Post-test productive	3.280 [*]	.536	.000	1.57	4.99	
	Combining Post-test productive	4.280 [*]	.536	.000	2.57	5.99	
	Control Post-test productive	5.720 [*]	.536	.000	4.01	7.43	

Table 8 shows that, on the immediate post-test, writing significantly did better than all groups. Moreover, writing group outperformed the combining and control groups. Post hoc tests also showed that the writing group obtained significantly higher scores than all groups on the delayed post-test.

DISCUSSION

According to the results of the study, the involvement load of the tasks has a positive effect on vocabulary learning and retention and on immediate and delayed post-tests. (Laufer, B., 2003) provided indirect evidence for the superiority of composing on long-term word retention. In two separate experiments she compared sentence writing (Experiment 1) and composing (Experiment 2) to reading comprehension with marginal glosses. Comparing across the two experiments, she concluded that the mean scores of the sentence and composition writing tasks were identical at the immediate test, but two weeks later the mean score for the composition task was higher than that of the sentence writing task. Therefore, it can be concluded that writing tasks can reinforce and encourage EFL learners to extend their knowledge of learning new lexical items (McDonough & Fuentes, 2015; Zou, 2017). In the delayed productive test, the writing task was also superior to other tasks in the retention of the target words.

The findings of the present study are in line with those obtained by Rott, S. (2005), Hulstijn, J.H. (1992), Shiping, D., Y. Chensong (2004), Hulstijn, J.H., P. Trompeter (1998), Keating, G.D. (2008), Laufer and Hulstijn (2001), Williams' (2012), McDonough & Fuentes (2015); Zou (2017). However, the findings of this study are not in line with some previous studies supporting the ILH (e.g., Bao, 2015; Keating, 2008; Marmol & Sanchez-Lafunte, 2013; Tahmasbi & Farvardin, 2017).

The research that has been done by (Shiping, D., Y. Chensong, 2004) also partially confirms the results of the current study. In their research, the construct of task-induced involvement was employed to design multiple word annotations and one word annotation. The result of the study revealed that both of these tasks were conducive to incidental vocabulary acquisition, but multiple word annotations were more beneficial than the latter one. For it requires more degrees of the involvement. The outcomes of (Keating, G.D., 2008) research also are in line with the outcomes of the current study. He checked out task effectiveness and word learning in second language reading comprehension domain. He observed that writing sentence tasks were more efficient than reading comprehension plus fill-in tasks and reading comprehension with marginal glosses. Therefore, he specified that tasks with higher involvement lead to greater gain in both passive and active word knowledge.

The results revealed that, regardless of vocabulary knowledge type, all output tasks were significantly better than the control group in receptive vocabulary knowledge test. Moreover, the findings showed that writing task was more effective than all groups.

In the immediate receptive vocabulary test, the writing task was better than other tasks, implying that the higher the involvement load a task has, the more effective the task will be for L2 vocabulary learning. Regarding productive vocabulary knowledge, writing was better than other tasks on both immediate and delayed productive tests.

CONCLUSION

The main focus of this study was on improving EFL learners' vocabulary acquisition through the involvement load hypothesis. The results of this study confirm the predictions of ILH that tasks with a higher involvement load better help EFL learners recall and retain the target words. The findings have also important implications for EFL practitioners. Teachers can arrange activities to help students develop their vocabulary learning through tasks with high involvement load and strong evaluation. Moreover, the teachers should apply useful opportunities to involve learners to learn more vocabulary knowledge and obtain meaningful usage of unknown words through tasks. Consequently, it is reasonable to conclude that retention of unfamiliar words is claimed to be conditional to the amount of involvement while processing new target words. Therefore, tasks with different involvement load will lead to different incidental acquisition (Laufer, B., J.H. Hulstijn, 2001). As well, the results of the study confirmed the validity of the involvement load hypothesis which makes it possible to translate and operationalize general cognitive notions of depth of processing and elaboration in terms of second language vocabulary learning tasks. According to (Craik, F.I.M., R.S. Lockhart, 1972), depth of processing theory claims that remembering information depends not only on having attended to it during its occurrence, but also on how deeply it is processed suggested that this notion is applicable in incidental vocabulary learning through giving rise to the involvement load hypothesis.

The findings of the present study are subject to a number of limitations. First, this study was conducted in Iran as an EFL context. Hence, future studies can examine the effectiveness of different vocabulary learning tasks in ESL and other EFL contexts. Second, the effect of each output task was studied separately. Thus, researchers can integrate the output tasks and investigate their effect on EFL learners' productive and receptive vocabulary knowledge. Third, this study explored only four vocabulary learning tasks. However, other types of vocabulary learning tasks and factors in EFL vocabulary learning were not examined. Therefore, further studies are needed to investigate other influential tasks and factors in L2 vocabulary learning.

REFERENCES

- Allan, D. (2004). *Oxford placement test*. Oxford, UK: Oxford University Press.
- Bao, G. (2015). Task Type Effects on English as a Foreign Language Learners' Acquisition of Receptive and Productive Vocabulary Knowledge. *System*, 53, 84-95.
- Bromberg, M., Lieb, J., & Traiger, A. (2011). *504 absolutely essential words*. Hauppauge, NY: Barron's Educational Series.
- Bruton, A. (2007). Vocabulary Learning from Dictionary Reference in Collaborative EFL Translational Writing. *System*, 35, 353e367.
- Eckerth, J., & Tavakoli, P. (2012). The Effects of Word Exposure Frequency and Elaboration of Word Processing on Incidental L2 Vocabulary Acquisition Through Reading. *Language Teaching Research*, 16(2), 227e252.

- Feng, T. (2014). Involvement Load in Translation Tasks and EFL Vocabulary Learning. *The New English Teacher*, 9(1), 83-101.
- Folse, K. S. (2006). The Effect of Type of Written Exercise on L2 Vocabulary Retention. *TESOL Quarterly*, 40, 273-293.
- Huang, L.-L., & Lin, C.-C. (2014). Three Approaches to Glossing and Their Effects on Vocabulary Learning. *System*, 44, 127e136.
- Hulstijn, J., & Laufer, B. (2001). Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning*, 51(3), 539e558.
- Keating, G. D. (2008). Task Effectiveness and Word Learning in a Second Language: The Involvement Load Hypothesis on Trial. *Language Teaching Research*, 12, 365-386.
- Kim, Y. (2008). The Role of Task-Induced Involvement and Learner Proficiency in L2 Vocabulary Acquisition. *Language Learning*, 58, 285-325.
- Laufer, B., & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-Induced Involvement. *Applied Linguistics*, 22(1), 1-26.
- Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental Vocabulary Acquisition: The Effects of Task Type, Word Occurrence and Their Combination. *Language Teaching Research*, 15(4), 391-411.
- Laufer, B., & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-Induced Involvement. *Applied Linguistics*, 22, 1-26.
- Lu, G.-S. (1994). *The English-Chinese dictionary*. Shanghai: Shanghai Translation Publishing House.
- Marmol, G. A., & Sanchez-Lafunte, A. A. (2013). The Involvement Load Hypothesis: Its Effect on Vocabulary Learning in Primary Education. *Revista Española de Lingüística Aplicada*, 26, 11-24.
- Martínez-Fernández, A. (2008). Revisiting the involvement load hypothesis: awareness, type of task and type of item. In M. Bowles, R. Foote, S. Perpinan, & R. Bhatt (Eds.), *Selected proceedings of the 2007 Second Language Research Forum* (pp. 210e228). Somerville, MA: Cascadia Proceedings Project.
- McCarty, M. (2005). *Discourse analysis for language teachers*. Cambridge, UK: Cambridge University Press.
- McDonough, K., & Fuentes, C. G. (2015). The Effect of Writing Task and Task Conditions on Colombian EFL Learners' Language Use. *TESL Canada Journal*, 32, 67-79.
- Min, H. T. (2008). EFL Vocabulary Acquisition and Retention: Reading Plus Vocabulary Enhancement Activities and Narrow Reading. *Language Learning*, 58, 73-115.
- Nation, P. (2003). Vocabulary. In D. Nunan (Ed.), *Practical English language teaching* (pp. 129-152). New York, NY: McGraw-Hill.
- Nation, P., & Meara, P. (2002). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 35-54). London, England: Arnold.

- Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.
- Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS version 15*. New York, NY: McGraw- Hill.
- Pichette, F., de Serres, L., & Lafontaine, M. (2012). Sentence Reading and Writing for Second Language Vocabulary Acquisition. *Applied Linguistics*, 33(1), 66e82.
- Plonsky, L., & Oswald, F. L. (2014). How Big is “Big”? Interpreting Effect Sizes in L2 Research. *Language Learning*, 64, 878-912.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Tang, C., & Treffers-Daller, J. (2016). Assessing incidental vocabulary learning by Chinese EFL learners: Testing the Involvement Load Hypothesis. In G. Yu & Y. Yin (Eds.), *Assessing Chinese learners of English: Language constructs, consequences and conundrums* (pp. 121-148). London, UK: Palgrave.
- Vidal, K. (2011). A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition. *Language Learning*, 61(1), 219e258.
- Walter, E., Woodford, K., & Good, M. (Eds.). (2008). *Cambridge advanced learner's dictionary* (3rd ed.). Cambridge, UK: Cambridge University Press.
- Webb, S. (2005). Receptive and Productive Vocabulary Learning: The Effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition*, 27, 33-52.
- Webb, S. (2005). Receptive and Productive Vocabulary Learning: the Effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition*, 27(1), 33e52.
- Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. Wilcox, R. R. (2012). Introduction to robust estimation and hypothesis testing. San Diego, CA: Elsevier.
- Wilcox, R. R., & Tian, T. S. (2011). Measuring Effect Size: a Robust Heteroscedastic Approach for Two or More Groups. *Journal of Applied Statistics*, 38(7), 1359e1368. *TESOL Quarterly*, 43(1), 55e77.
- Williams, J. (2012). The Potential Role(s) of Writing in Second Language Development. *Journal of Second Language Writing*, 21, 321-331.
- Yaqubi, B., Rayati, R. A., & Allemzade Gorgi, N. (2010). The Involvement Load Hypothesis and Vocabulary Learning: The Effect of Task Types and Involvement Index on L2 Vocabulary Acquisition. *Journal of Teaching Language Skills*, 29, 145-163.
- Zou, D. (2017). Vocabulary Acquisition Through Cloze Exercises, Sentence-Writing and Composition-Writing: Extending the Evaluation Component of the Involvement Load Hypothesis. *Language Teaching Research*, 21, 54-75.