The Role of Task Type in L2 Vocabulary Acquisition: A Case of Involvement Load Hypothesis

Dr. Abdullah Sarani

Assistant Professor, University of Sistan and Baluchestan, Zahedan, Iran

E-mail: sarani_ling@hamoon.usb.ac.ir

Dr. GitiMousapourNegari

Assistant Professor, University of Sistan and Baluchestan, Zahedan, Iran

E-mail: mousapour@hamoon.usb.ac.ir

MassoumehGhaviniat

M.A. Student of TEFL, University of Sistan and Baluchestan, Zahedan, Iran

E-mail: m.ghaviniat@gmail.com

Abstract

Based on Involvement Load Hypothesis (Laufer & Hulstijn, 2001), this study aimed to examine the effect of involvement load, and task type on vocabulary acquisition. Six classes of EFL learners were assigned to one of six experimental groups with different involvement loads, thus leaving three groups with receptive tasks, and three ones with productive tasks. Learners read a text and completed 10 vocabulary tasks focused on the target words while time on task was controlled across groups. The knowledge of target words was tested in two posttests. Predictably, the findings indicated that tasks with higher involvement loads were more effective for vocabulary learning than tasks with lower involvement loads. We also compared receptive tasks with productive ones of the same load condition. Contrary to Involvement Load Hypothesis, productive tasks were more effective than receptive ones. We also found that the time on task does not any effect on task efficacy.

Key words: involvement index, receptive tasks, productive tasks

Introduction

There is a general agreement among most researchers that the vocabulary is one of the main constituents of a language, and acquiring L2 vocabulary is the prerequisite of second language learning. In fact, this is the vocabulary by which L2 learners will be able to do each of four skills (Richards & Renandya, 2002). The importance of vocabulary knowledge in academic situations

(Donley & Reppen, 2001), in reading ability (Kitajima, 2001; Meara & Fitzpatrick, 2000), and in human communications (Coady & Huckin, 1997) has been greatly emphasized. About the number of L2 words needs to be learned, some researchers proposed that 5,000 words is a least lexical necessity for a L2 learners of English to understand general, non-specialized (Nation, 1990; Laufer, 1997) or unsimplified (Hirsh & Nation, 1992) texts. However, for understanding the specialized and academic texts, 7,000 (Groot, 1994, cited in Groot, 2000) or 10,000 (Schmitt, 2000) word stock is required. Similarly, 5,000 words are prerequisite for communicative skills in a second or foreign language (Nation, 1993, cited in Prince, 1996). Accordingly, the first step for many foreign or L2 learners is to grab and memorize a large stock of vocabulary. But the question is how?

The accepted view among most researchers (e.g., Swanborn & de Glopper, 1999; Nagy & Herman, 1987; etc.) is that it is not possible for L2 learners to learn such a large stock of vocabulary only through explicit instruction of vocabulary because according to Schmitt (2000), it is very time- consuming and too laborious. In addition, as Krashen (1989, cited in Keating, 2008) states, the majority of word learning by L2learners occurs incidentally. Shortly after developing the incidental vocabulary learning hypothesis by Nagy and Herman (1985), a wide variety of studies were performed in order to discover the most effective factors on incidental word learning during different kinds of tasks. The results of studies done on this area revealed an extensive diversity of factors which were effective in promoting incidental word learning. For example, one type of research related to incidental vocabulary learning put emphasis mainly on learner factors (Prince, 1996; Swanborn & de Glopper, 2002). On the other hand, while some studies examined the impact of contextual cues such as marginal glosses (Hulstijn, 1992; Watanabe, 1997), others considered the use of dictionary (Luppescu & Day, 1993; Knight, 1994) as an issue affecting incidental vocabulary learning. Another type of studies investigated the effects of text-based (Joe, 1995, 1998), word-focused (Wesche & Paribakht, 2000; Laufer, 2001; Paribakht & Wesche, 1997), and interactional tasks (Newton, 1995; Loschky, 1994; Ellis, 1995) on incidental vocabulary learning. Additionally, the use of negotiation and interaction (Ellis & He, 1999; de la Fuente, 2002) was also considered as an effective factor on incidental learning of L2 vocabulary.

In each of these studies, one task was superior to another in term of incidental vocabulary learning. In explaining this superiority, most authors indicated that the more effective task requires a "deeper level of processing" (Craik & Lockhart, 1972) than the other task.

Nonetheless, Craik and Lockhart's (1972) depth of processing has been criticized for not having a clear-cut and simple definition about different levels of processing (Craik & Tulving, 1975; Nelson, 1977; Eysenck, 1978; Baddeley, 1999; Laufer & Hulstijn, 2001). Accordingly, in order to provide a more clear-cut definition of depth of processing, the Involvement Load Hypothesis was formulated by Laufer and Hulstijn (2001).

The Involvement Load Hypothesis

The Involvement Load Hypothesis is an incidental vocabulary learning theory that formulated the criteria which explain why some specific tasks lead to better vocabulary retention than others. In their hypothesis, they proposed the construct of task-induced involvement load which calculates the amount of task efficacy in the retention of new L2 vocabulary in an incidental condition. This construct is comprised of three principal components: need, search, and evaluation. The need component refers to whether for completing a task the learner is supposed to know the meaning of the new words. Two levels of importance for need were offered: moderate and strong. Need is moderate when it is externally enforced by the teacher or the task, and strong when it is intrinsically imposed by the learner.

Search, as opposed to need, signifies the endeavor of discovering the meaning of a new L2 word or discovering the L2 form of a word in L1. Unlike need, search can be present or absent. It is present while the learner's endeavor of discovering the meaning of unfamiliar words to complete a task exist; however, it is absent while such an endeavor does not exist. Evaluation entails reaching a conclusion about the meaning of a word during tasks, for example, "a comparison of a given word with other words , a specific meaning of a word with its other meanings, or combining the word with other words in order to assess whether a word does or does not fit its context " (Laufer & Hulstijn, 2001, p. 14). Like need, evaluation can also be moderate or strong. Evaluation is moderate while the learners are required to compare several lexical items with each other (as in matching tasks), or compare different meanings of a lexical item in a provided text (as in a homonym). However, strong evaluation makes learners to combine new lexical items, and create a novel sentence.

Combining all the three factors with their levels of importance in a task makes the task-induced involvement load. Laufer and Hulstijn (2001) declared that tasks with   higher involvement loads promote better vocabulary retention than tasks with lower involvement loads. But how can we determine one task's involvement load in a numerical fashion? Accordingly, in order to compare different tasks with each other in a numerical fashion, Hulstijn and Laufer (2001)

offered the involvement index, which appoints numerical weights in which "absence of a factor is marked as 0, a moderate presence of a factor as 1, and strong presence as 2" (p. 544). Therefore, each task can have of an involvement index of 0 (lowest index) to 5 (highest index). In their hypothesis, Laufer and Hulstijn (2001) declared that any special task type (e.g., output) does not consider more effective than other type (e.g., input). They asserted that this is just the level of involvement load of a task which determine task's efficacy. In other words, they stated that two input and output tasks with the same load conditions will act equally on vocabulary acquisition. So, the equality of the involvement loads for different tasks types (e.g., input vs. output) calls for further research.

Empirical Studies on the Involvement Load Hypothesis

Rightly after developing the Involvement Load Hypothesis, Hulstijn and Laufer (2001) investigated the effect of involvement load on short- and long-term retention of 10 unfamiliar words by advanced EFL learners in two different experiments. They compared three learning tasks with varying involvement loads: reading comprehension with marginal glosses (index = 1), reading comprehension plus fill- in (index = 2), and writing a composition and using the target words (index = 3). Immediately after treatment, the learners were asked to write L1 translations or L2 definitions for the 10 target words in order to measure their short-term retention of target words. The same posttest was again administered 1 or 2 weeks later in order to measure student's long-term retention. The results of the Hebrew-English Experiment which provided strong support for the Involvement Load Hypothesis revealed that, on both posttests, the composition group scored significantly higher than the fill- in group, and the fill- in group scored significantly higher than the reading group. Nonetheless, the results of the Dutch- English Experiment which provided partial support for the hypothesis reported that, on both posttests, the composition group performed significantly better than the fill-in and the reading groups; yet, the fill-in group did not performed significantly better than the reading group.

By stating the limitations of Hulstijn and Laufer's (2001) study, Keating (2008) investigated whether the low-proficiency learners can also benefit from the more involving tasks, and whether the learners can gain the same word knowledge on passive and active tests. In order to reach a certainty about these questions, the low- proficiency learners of Spanish randomly completed one of the three tasks with different involvement loads: reading comprehension with marginal glosses (index = 1), reading comprehension plus fill-in (index = 2), and writing original sentences by using target words (index = 3). After task completion and two weeks later, the

learners' knowledge of target words was assessed through two passive and active tests. Partially confirming the Involvement Load Hypothesis, the results of both immediate and delayed passive tests reported that Task 2 and 3 resulted in higher retention scores comptared to Task 1. However, Task 3 was not more effective than Task 2. On the other hand, the results of immediate active test which firmly supported the hypothesis revealed that learners in Task 2 and 3 promoted better word retention than those in Task 1, and learners in Task 3 also performed better than those in Task 2. In the delayed active test, however, learners in Task 3 did not performed very better than those in Task 1 or Task 2. In short, Keating's (2008) study claimed that the Involvement Load Hypothesis can be generalized to low proficiency learners, and it can also affects similarly on the learner's passive and active word knowledge.

Following Keating (2008), Kim (2011) tested the Involvement Load Hypothesis in an ESL setting, across different task types and proficiency levels with a controlled time on task. In his first experiment, he tested the efficacy of three tasks with varying involvement loads within two different proficiency levels. In each proficiency level, learners randomly completed one of three tasks: Reading, Gap-fill, and Composition. In order to assess L2 learners' initial learning and retention of target words, two immediate and delayed posttests were conducted. The results of both posttests showed that the Composition group (index = 3) yielded significantly higher scores than the Reading (index = 1) and Gap-fill (index = 2) groups. However, the participants in Gap-fill group gained significantly better scores than those in the Reading group just on the delayed posttest. In a nutshell, the results of immediate posttest partially supported the Involvement Load Hypothesis while those of delayed posttest fully supported this hypothesis. In his second experiment, Kim (2011) investigated whether two tasks with equal involvement loads affect similarly on learning of target words. To meet this, he compared the writing composition (index = 3) with the writing sentence task (index = 3). The results of both posttests revealed that these two tasks with identical involvement loads affected equally on the initial learning and retention of target words across two different proficiency levels, a claim which was supported by the Involvement Load Hypothesis.

By considering the research studies of the Involvement Load Hypothesis, it appears that time on task has not been well considered in these studies. Initially, Folse (2006) claimed that the efficacy of one task over another might be due to the length of time needed for task completion. Similarly, Keating (2008) argued that when time on task was taken into account, the benefits connected to more involving tasks faded. As opposed to Folse (2006) and Keating (2008), Kim

(2011) empirically tested the Involvement Load Hypothesis with a controlled time on task across groups, and proved that the results of his study were in pattern with the predictions of the Involvement Load Hypothesis. Due to these contradictory results about the role of time on task, we are still in need of further research to test this hypothesis with a controlled time on task from the outset of the study.

Further research is also needed to investigate the underlying assumptions of the Involvement Load Hypothesis. For example, Laufer and Hulstijn (2001) claimed that any particular task type— be it input or output—is not considered superior or more effective, and that the only influential factor in task efficacy is the task's level of involvement load. Consequently, more research is needed to examine whether tasks with similar levels of involvement load but from different types—input vs. output—will have similar effects on vocabulary acquisition. To meet these two purposes, the researcher designed three receptive and three productive vocabulary tasks with varying involvement loads. In the light of the purposes of the study, the following research questions were posed:

1. Given English receptive vocabulary tasks, will Iranian EFL learners obtain better initial learning and retention of new vocabulary in higher task load conditions compared to lower ones? If so, will the benefits of tasks hold up over time?

2. Given English productive vocabulary tasks, will Iranian EFL learners obtain better initial learning and retention of new vocabulary in higher task load conditions compared to lower ones? If so, will the benefits of tasks hold up over time?

3. Given English receptive and productive vocabulary tasks with the same levels of involvement index, will Iranian EFL learners obtain the same initial learning and retention of new vocabulary on both types of tasks?

Methodology

Participants

Six intact classes of second-year English major university students homogenized based on the TOEFL exam were selected for this study. All of them were EFL learners, and their first language was Persian. Initially, 179 students took part in this. However, all of them were not present on the delayed posttest because they were not informed of the delayed posttest due to the incidental learning nature of this study. Moreover, the data from two subjects was excluded from the study because of having the knowledge of more than two target words. Accordingly, the final number of students taking part in this study was 162. Each of the six intact classes was randomly

assigned to one of the six experimental groups, thus leaving three groups with receptive tasks, and the other three ones with productive tasks.

Materials

The Target Words

The 10 target words which were supposed to be unfamiliar to the learners were chosen for examination from the reading text "Coping with Procrastination" from Kim's (2011) study. To make sure that the text would be of an appropriate level for our participants, the length and complexity of the text were modified by the researcher and two experienced English teachers. The unfamiliarity of target words was checked through the pilot study with a group of participants who would not participate in the experiment. These participants, who had the same proficiency level of ours, were given a list of 10 target words and asked to translate them. The overall mean score was .2 out of 10 target words. Thus, the target words were unfamiliar within this proficiency level. However, as a final confirmation, the preknowledge of the participants in the main study was also checked in the immediate posttest .The ten target words chosen from the text are: apprehensive, oration, vexed, spawn, envision, abate, caveat, assiduous, stymie, and divulge. It should be mentioned that the 10 target words were emphasized by printing them in a bold face, and glossed in L1 (Persian) as well as L2 (English) in the margin of the text.

The Graphic Organizers

A set of graphic organizers designed by Kim (2011) were also taken and modified according to the revised text. We used the graphic organizers, which acted like a comprehension activity, in order to control the time on task across the six groups. So, the participants in True-false, Matching, and Fill-in the blank task conditions were asked to answer the graphic organizers because they took less time than the other groups, as it was discovered in the pilot study.

Vocabulary Task Conditions

To address the first research questions, the researcher designed three receptive vocabulary tasks with varying involvement loads: True-false, involvement = 1; Matching, involvement = 2; and Multiple-choice, involvement = 3.

True-False Task Condition

Participants assigned to the True-false task condition were asked to read the marginally glossed text and next complete the graphic organizers to control for time on task across all groups. Afterwards, they were given the 10 True-false vocabulary tasks focused on the target words. In terms of the Involvement Load Hypothesis, this task induced a moderate need (the knowledge of

target words was relevant to answering the tasks), but neither search nor evaluation. Its involvement index was thus 1 (1 + 0 + 0).

Matching Task Condition

Participants in the Matching task condition were also asked to read the text and complete the graphic organizers. After that, they were given 10 Matching vocabulary tasks focused on the target words. This task induced moderate need, no search, and moderate evaluation. Evaluation was moderate because the participants had to distinguish among different definitions to answer the Matching vocabulary tasks. Therefore, the involvement index of this task was 2 (1 + 0 + 1).

Multiple-Choice Task Condition

Participants in the Multiple-choice task condition were provided with the same text in the last two groups; however, the text was not marginally glossed. The participants' task was to read the text by looking up the target words in a dictionary; afterwards, they were given 10 Multiple-choice vocabulary tasks focused on the target words. This task induced moderate need and moderate evaluation (because four options in each of the Multiple-choice vocabulary tasks must be assessed against each other). The search factor was also present here. Therefore, the involvement index of this task was 3 (1 + 1 + 1).

To address the second research questions, the researcher designed three productive vocabulary tasks with different involvement loads: Short-response, involvement = 1; Fill-in the blank, involvement = 2; and Sentence writing, involvement = 3.

Short-Response Task Condition

Participants in the Short-response task condition received the same marginally glossed text to read, and then to complete the 10 Short-response vocabulary tasks focused on the target words. The involvement index of this task was 1 (1 + 0 + 0). Need was moderate, but search and evaluation were absent.

Fill-in the Blank Task Condition

Participants performing the Fill-in the blank task condition were asked to read the same text and then complete the graphic organizers. Afterwards, the learners were required to complete the 10 Fill-in the blank vocabulary tasks focused on the target words with the most suitable word from 15 glossed words—10 target words plus 5 additional words— in the reading text. Consequently, the students could not narrow down the choices as they progress through the vocabulary tasks simply by omitting the words that they have already used .This task induced moderate need, no

search, and moderate evaluation. The evaluation was moderate because the 15 glossed words must be assessed against each other. Thus, the involvement index of this task was 2 (1 + 0 + 1).

Sentence Writing Task Condition

Participants in the Sentence writing task condition received the same marginally glossed text, and were asked to read the text. Then, they were required to write L2 (English) sentences by using the 10 target words. The involvement index of this task was 3 (1 + 0 + 2), that is, a moderate need, no search, and a strong evaluation. The evaluation was strong because the participants were required to assess the target words within appropriate collocations in order to generate a new context.

Vocabulary tests

The present study administered two immediate and delayed posttests to respectively assess the learners' initial learning and retention of target words. Upon the completion of tasks, and three weeks later, we unexpectedly tested the participants' knowledge of target words through a modified version of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1997) in all six task conditions (see Figure 1).

Self-report Categories	
I	I don't remember having seen this word before.
II	I have seen this word before, but I don't know what it means.
III	I have seen this word before, and I think it means ------------------. (synonym or translation)
IV	I can use this word in a sentence: ------------------. (write a sentence.) *(If you do this section, please also do Section III.)*

Figure 1. Modified VKS self-report categories

The scoring procedure of the modified VKS can be presented in this way:

- 1 point: The word is not familiar at all (category I).
- 2 points: The word is familiar but its meaning not known (category II).
- 3 points: A correct synonym or translation is given (category III).
- 4 points: The word is used with semantic appropriateness in a sentence (category IV).
- 5 points: The word is used with semantic appropriateness and grammatical accuracy in a sentence (category IV).

9

It should be noted that wrong responses in self-report categories III or IV will lead to a score of 2. Overall, the possible test score for both posttests was 10–50. On both posttests, we gave the learners the 10 target words in the form of VKS, and asked them to complete it. The learners were also asked to point out if any of the words were familiar to them before doing the task. If a learner was previously familiar with more than two words, the data collected from that learner was removed from the analysis.

Procedure

This research was accomplished on two separate days. In the first day, we administered the treatment and the immediate posttest. But, the delayed posttest was carried out three weeks later. On the treatment day, each of the six intact classes was randomly asked to complete one of the following task conditions: True-false, Matching, Multiple-choice, Short-response, Fill-in the blank, or Sentence writing. In each group, the participants were asked to read the text and complete the 10 vocabulary tasks. To control the time on task, we also added a set of graphic organizers to the True-false, Matching, and Fill-in the blank groups. Each of six task conditions took 50 minutes to complete. Due to the nature of the study, incidental learning, the participants were not informed of the upcoming immediate or delayed posttests because according to Hulstijn (2001), test announcement is an indication of intentional word learning. Accordingly, after task completion, and three weeks later, the participants were unexpectedly given the immediate and delayed posttest in a modified form of VKS in order to respectively measure the initial learning and retention of target words.

Data Analysis

The first two research questions were posed to assess if the level of involvement load affects the initial learning and retention of new vocabulary when tasks with different involvement loads are administered. The dependent variable for these two questions was the scores of the immediate and delayed posttests, and the independent variable was the level of involvement load. In order to examine the impact of the independent variable on the dependent variable, the VKS scores of both posttests were submitted to four, one-way ANOVAs. The Scheffe post hoc contrasts were then computed to locate significant differences among pairs. Additionally, six paired samples t-tests were performed to further investigate if the benefits of tasks will hold up over time. Unlike the first two, the third research question examined if the type of vocabulary task affects the initial learning and the retention of new words when two different types of task (receptive or productive) with the same involvement loads are administered. The dependent variable in this

question was the scores of both posttests, and the independent variable was the type of vocabulary task at two levels: receptive and productive. Six independent samples t-tests were performed to compare the receptive tasks with the productive ones of the same load condition. The alpha level was set at .05 when significant results were found.

Results

Data Analysis of Three Receptive Tasks

The descriptive statistics of the three receptive vocabulary tasks in Table 1 demonstrate that, on both posttests, the Multiple-choice group performed better than the Matching group, which, in turn, performed better than the True-false group. To determine if these differences were statistically significant, the scores of each posttest were then submitted to a one-way ANOVA.

Table 1. Descriptive Statistics of the Immediate and Delayed Posttests
for the Three Receptive Vocabulary Tasks

	N	Mean	Std. Deviation	Std. Error
Immediate Posttest				
True-false (1)	29	17.06	1.386	.257
Matching (2)	27	20.03	1.698	.326
Multiple-choice (3)	27	27.92	3.395	.653
Total	83	21.56	5.144	.564
Delayed Posttest				
True-false (1)	29	14.34	1.421	.263
Matching (2)	27	17.33	1.901	.365
Multiple-choice (3)	27	21.07	1.206	.232
Total	83	17.50	3.167	.347

Note. The involvement index for each task is indicated in parentheses. The possible VKS scores in all three vocabulary tasks ranged from 10 to 50.

The results of both ANOVAs revealed a main effect for the level of task's involvement load for both the immediate, $F (2, 80) = 162.519$, $p< .001$, and the delayed posttest, $F (2, 80) = 134.678$, $p< .001$. In fact, there is a significant difference among the tasks with different levels of involvement load on both posttests. The results of two Scheffe *post hoc* tests also indicated that the Multiple-choice group significantly outscored both the Matching and the True-false groups, and the Matching group also significantly outscored the True-false group.

Comparing the means of the immediate with those of the delayed posttest for each of the three receptive vocabulary tasks, the results of three paired samples *t*-tests revealed that there was a significant decrease in the mean scores of the delayed posttest for all the three receptive

11

vocabulary tasks, that is, for the True-false task [t (28) = 11.471, p< .001], for the Matching task, [t (26) = 8.980, p< .001], and for the Multiple-choice task, [t (26) = 12.486, p< .001]. In addition to the t-tests results, the sharp decline of the lines in Figure 2 also showed that the performance in all the three groups degenerated significantly on the delayed posttest.
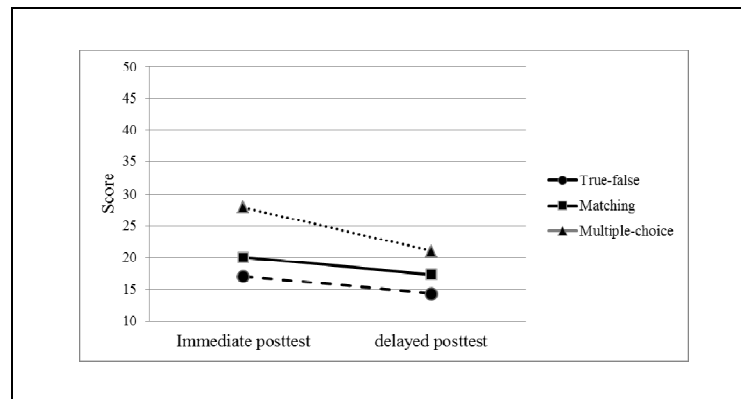

Figure 2. The scores of the immediate and delayed
posttests for the three receptive vocabulary tasks

Data Analysis of Three Productive Tasks

The descriptive statistics of the three productive vocabulary tasks in Table 2 suggest that the mean score of the Sentence writing group was higher than those of the Fill-in the blank and the Short-response groups on both posttest; however, there was not a much difference between the mean scores of the latter two groups on the delayed posttest.

Table 2. Descriptive Statistics of the Immediate and Delayed Posttests
for the Three Productive Vocabulary Tasks

	N	Mean	Std. Deviation	Std. Error
Immediate Posttest				
Short-response (1)	26	21.88	1.728	.338
Fill-in the blank (2)	28	24.75	3.038	.574
Sentence writing (3)	25	28.56	2.328	.465
Total	79	25.01	3.625	.407
Delayed Posttest				
Short-response (1)	26	17.34	1.719	.337
Fill-in the blank (2)	28	18.10	.875	.165
Sentence writing (3)	25	21.68	1.625	.325
Total	79	18.98	2.350	.264

Note. The involvement index for each task is indicated in parentheses. The possible VKS scores in all three vocabulary tasks ranged from 10 to 50.

To determine the statistical differences among groups, two one-way ANOVAs were conducted. The ANOVA results indicated that significant differences were found among the three productive vocabulary tasks on both the immediate, F (2, 76) = 47.780, p< .001, and the delayed posttest, F (2, 76) = 65.653, p< .001. The results of Scheffe tests also demonstrated that the Sentence writing group performed significantly better than the Fill-in the blank and the Short-response groups on both posttests, but the Fill-in the blank group performed significantly better than the Short-response group only on the immediate posttest.

Regarding the means of the immediate and delayed posttests for each of the three productive vocabulary tasks, the t-tests results revealed a significant decrease in the mean score of the delayed posttest for the Short-response [t (25) = 11.781, p< .001], for the Fill-in the blank [t (27) = 11.588, p< .001], and for the Sentence writing group [t (24) = 14.975, p< .001]. In addition to the t-tests results, the downward lines in Figure 3 also revealed a general degeneration of the performance in all the three groups on the delayed posttest, suggesting that the interval between the two posttests may be a main reason for the decline in performance of all groups.
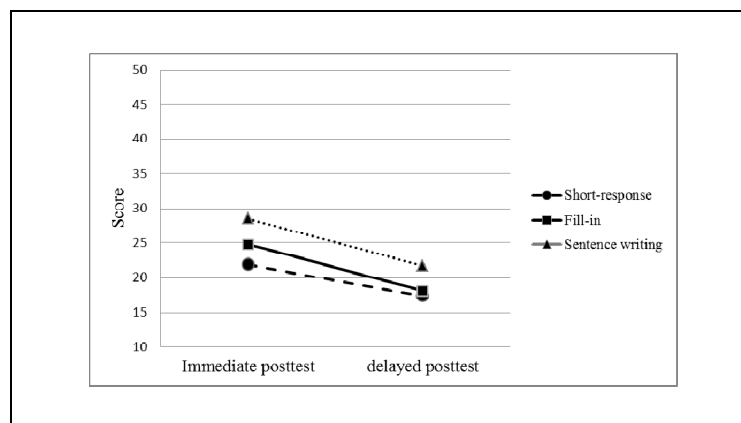

Figure 3. The scores of the immediate and delayed posttests for the three productive vocabulary tasks

The Comparison between Receptive and Productive Tasks

Regarding the comparison between the true-false (load = 1) and the short-response group (load = 1), the results of two independent t-tests showed a significantly better performance for the Short-response group on both the immediate, t (53) = -11.450, p< .001, and the delayed posttest, t (53) = -7.084, p < .001. In the case of comparison between the matching (load = 2) and the fill-in the blank group (load = 2), the results of t-tests revealed that the Fill-in the blank group performed significantly better than the Matching group on the immediate posttest, t (42) = -7.134, p< .001; however, this preference of the Fill-in the blank group was not observed on the delayed posttest,

t (36) = -1.927, *p* = .062 > .05. Unlike the last two pairs, the *t*-tests' results of the comparison between the multiple-choice (load = 3) and the sentence writing group (load = 3) revealed that there was not a significant difference between these two groups on both the immediate, *t* (50) = -.779, *p* = .440> .05, and the delayed posttest, *t* (50) = -1.534, *p* = .131 > .05.

Discussion and Final Considerations

The first two research questions were framed to investigate whether tasks with a higher involvement load achieve better vocabulary scores than tasks with a lower involvement load while time on task is controlled across different groups. The results of the first research question on both posttests fully supported the Involvement Load Hypothesis in that the Multiple-choice group with the highest involvement load (3) produced better initial learning and retention of target words than the Matching group with the lower involvement load (2), which, in turn, performed better than the True-false group with the lowest involvement load (1). However, the results of the second research question partly supported the Involvement Load Hypothesis in that the Sentence writing group (involvement = 3) performed significantly better than the Short-response (involvement = 1) and the Fill-in the blank group (involvement = 2) on both posttests, but the Fill-in the blank group performed significantly better than the Short-response group only on the immediate, but not the delayed posttest.

Unlike the first two, the third research question was constructed to investigate Laufer and Hulstijn's (2001) claim that no particular task type— be it input or output—is considered superior or more effective, and that the only determining factor in task efficacy is the degree of involvement load that a task induces. To meet this end, the researcher compared the receptive tasks with the productive ones of the same load condition. Worded differently, she compared the True- false (load = 1) with the Short-response task (load = 1), the Matching (load = 2) with the Fill-in the blank task (load = 2), and the Multiple-choice (load = 3) with the Sentence writing task (load = 3). Contrary to the predictions of the Involvement Load Hypothesis, the results of the first pair comparison revealed the better performance of the Short-response (a productive task) over the True-false (a receptive task) on both posttests. Similarly, contra to the Involvement Load Hypothesis, the Fill -in the blank (a productive task) performed significantly better than the Matching (a receptive task) on the immediate posttest; however, this preference of the Fill-in the blank group was not observed on the delayed posttest. Unlike the last two pairs, the results of the third pair comparison completely fulfilled the predictions of the Involvement Load Hypothesis in

that the Sentence writing (a productive task) performed as well as the Multiple-choice (a receptive task) on both posttests.

Overall, the results of the first research question on both posttest, and the results of the second research question on the immediate posttest are in harmony with those obtained in Hulstijn and Laufer's (2001) Hebrew -English Experiment, Kim's (2011) first Experiment on the delayed posttest, and Keating's (2008) active word recall on the immediate posttest in that they all supported the Involvement Load Hypothesis. Similarly, the results of the second research question on the delayed posttest are exactly the same as those obtained in Hulstijn and Laufer's (2001) Dutch-English Experiment and Kim's (2011) first Experiment on the immediate posttest. One possible explanation for the observed results in this part is that the amount to which different levels (moderate and strong) of each factor (need, search, and evaluation) which constitute the overall involvement load might not be the same (Kim, 2011).

Nevertheless, the results of the third research question are considerably in conflict with the predictions of the Involvement Load Hypothesis. This hypothesis does not predict that any output task will lead to better results than any input task when they both have the same involvement load. On the contrary, we found that despite the involvement load induced by the task, the type of task is also effective in learning new words. In other words, two different types of tasks (receptive and productive) with the same level of involvement load may not have the same results in L2 vocabulary retention. This explanation provides support for Swain's (1985) Output Hypothesis which claims that the act of production demands deeper cognitive effort and can contribute more to word learning than the mere reading of a text which is an act of reception. In other words, what contribute to word retention is not merely the involvement load factors (i.e., need, search, and evaluation), rather other elements such as task type may be equally important, a claim which is in contrast with the Involvement Load Hypothesis.

In general, the findings of this study clearly run counter to some of the previous studies (e.g., Ellis, 1995; Folse, 2006; Webb, 2005; Hulstijn & Laufer, 2001; Keating, 2008) which claimed that controlling for time on task will diminish the effect of more involving tasks on vocabulary learning. However, like Kim (2011), we found that even if the time on task is controlled across different groups, the more involving tasks will perform better than the less involving ones in vocabulary scores. Apart from a significant task's involvement load effect, the results of the study also showed a significant decrease in the performance of all six groups on the delayed posttest, as it was observed in some of the previous studies (e.g., Watanabe, 1997; Hulstijn &

Laufer, 2001; Keating, 2008). This finding can be explained by Hulstijn's (2001) claim that a decrease in knowledge over time is natural in the absence of repetitive practice or additional exposure to the newly learned material.

REFERENCES

Baddeley, A. D. (1999). *Essentials of human memory*. Hove, UK: Psychology Press.

Coady, J. & Huckin, T. (eds.) (1997).*Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Craik, F. I. M. & Tulving, E. (1975).Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition, 24*, 81–112.

Donley, K. M. & Reppen, R. (2001).Using corpus tools to highlight academic vocabulary in SCLT.*TESOL Journal*, 10, 7-12.

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16* (4), 409-441.

Ellis, R. & He, X. (1999).The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*, 285–301.

Eysenck, M. W. (1978). Levels of processing: A critique. *British Journal of Psychology, 69*, 157-169.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention.*TESOL Quarterly, 40* (2), 273–293.

Groot, P.J.M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning and Technology, 4* (1), 60-81.

Hirsh, D. & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8* (2), 689–696.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Macmillan.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hulstijn, J. H. & Laufer, B. (2001).Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539–558.

Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research, 11* (2), 149–158.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19* (3), 357–377.

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.

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 61* (1), 100-140.

Kitajima, R. (2001). The effect of instructional conditions on students' vocabulary retention. *Foreign Language Annals, 34*(5), 470-482.

Knight, S. M. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal, 78*, 285–299.

Laufer, B. (1997). The lexical plight in second language reading: words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *second language vocabulary acquisition: A rational for pedagogy* (pp. 20-34). Cambridge: Cambridge University Press.

Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language. *Prospect, 16*(3), 44–54.

Laufer, B. & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: the construct of task induced involvement. *Applied Linguistics, 22*(1), 1-26.

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition, 16*, 303-323.

Luppescu, S. & Day, R. R. (1993).Reading, dictionaries, and vocabulary learning. *Language Learning, 43* (2), 263–287.

Meara, P. & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19-30.

Nagy, W. E. & Herman, P. A. (1985).Incidental vs. instructional approaches to increasing reading vocabulary. *Educational perspectives*, 23, 16-21.

Nagy, W. E. & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary acquisition,* (pp. 19-35). Hillsdale, NJ: Erlbaum Associates.

Nation, I. S. P. (1990).*Teaching and learning vocabulary*. Boston: Heinle & Heinle Publishers.

Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 151-171.

Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research, 11* (2), 159-177.

Paribakht, T. S. & Wesche, M.B. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp.174–200). Cambridge: Cambridge University Press.

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal, 80* (4), 478–493.

Richards, J.C. & Renandya, W.A. (2002).*Methodology in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-252). Rowley, MA: Newbury House.

Swanborn, M. S. L. & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of educational research*, 69 (3), 261-285.

Swanborn, M. S. L. & de Glopper, K. (2002).Impact of reading purpose on incidental word learning from context. *Language Learning, 52* (1), 95-117.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increasing processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19*, 287-307.

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.

Wesche, M. B. & Paribakht, T. S. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *The Modern Language Journal, 84* (2), 196-213.