The Impact of Skill Integration on Task Involvement Load

Atefeh Amini [1], Parviz Maftoon [2*]

1, 2. Department of English, Science and Research Branch, Islamic Azad University, Tehran, Iran
*Corresponding author: pmaftoon@srbiau.ac.ir

Abstract

The present study investigated whether word learning and retention in a second language are contingent upon a task's involvement load, i.e., the amount of need, search, and evaluation the task imposes. Laufer and Hulstijn (2001) contend that tasks with higher degrees of these three components induce higher involvement load, and are, therefore, more effective for word learning. To test this claim, 64 Iranian intermediate EFL learners were selected based on their performance on the Preliminary English Test (PET). The participants were randomly assigned to two equal groups. Each group completed different vocabulary learning tasks that varied in the amount of involvement they induced. The tasks were jigsaw task (Group A) and information gap task (Group B). During the ten treatment sessions, recall and retention of the 100 unfamiliar target words were tested through immediate and delayed posttest. Data were analyzed using repeated measure ANOVA. The results indicated that learners benefited more from jigsaw task with higher involvement load. This study supported the Involvement Load Hypothesis, suggesting that higher involvement induced by the task resulted in more effective recall; however, no significant difference was observed between the two tasks in the retention of the unknown words.

Keywords: involvement load, information gap task, jigsaw task, vocabulary learning

Introduction

It is generally acknowledged that vocabulary is one of the essential components of a language. Second language (L2) vocabulary learning is a complex process involving not only word meaning comprehension but also retention, retrieval, and use of words. Researchers believe that sufficient vocabulary knowledge is a prerequisite for reading comprehension (see Decarrico, 2001; Laufer, 1992; Nation, 1993; Paribakht & Wesche, 1997). Many studies emphasize the importance of vocabulary learning for L2 learners in speaking (Hincks, 2003; Joe, 1998), listening comprehension (Elley, 1989; Ellis, 1994), and writing (Hinkel, 2001; Laufer & Nation, 1995; Lee, 2003).

Vocabulary learning is a matter of considerable concern for both language teachers and L2 learners. In this regard, Nation (2001) states L2 learners soon discover that their lack of vocabulary knowledge impede their ability to comprehend and express themselves clearly. Therefore, one of the main difficulties for L2 learners is the vast number of words they need to know in order to communicate fluently. On the other hand, language learners often complain that they quickly forget newly acquired words. They usually look for effective ways to increase possibilities for saving new words in their long term memory, but forgetting is a common problem. Most teachers also know this problem but do not know how best to help their students. Thus, assisting students in vocabulary size expansion and new word retention are important issues which are needed to be taken into consideration.

Numerous current studies on vocabulary acquisition are based on a cognitive processing view of learning (Schmidt, 2001; Craik, 2002), of which the Involvement Load Hypothesis (ILH), as proposed by Laufer and Hulstijn (2001), has drawn researchers' attention. Based on ILH, word learning and retention are dependent upon the amount of mental effort or involvement that a task imposes. According authors, the hypothesis does not predict that any output task will lead to better results than any input task. It predicts that higher involvement in a word induced by the task will result in better retention regardless of whether it is an input or an output task.

ILH consists of three basic components--need, search, and evaluation--each with two degrees of prominence (moderate and strong). The need component is "the motivational, non-cognitive dimension of involvement" (Laufer&Hulstijn, 2001, p. 14). According to the authors, need is moderate when it is imposed

externally (e.g., teacher asks to use a word in a sentence). Need is strong when it is intrinsically motivated that is, self-imposed by the learner (e.g., learner's decision to look up a word in a dictionary).

Search and evaluation are the two cognitive dimensions of involvement. Search is the attempt to find the meaning of an unknown L2 word by consulting a dictionary or another authority (e.g., trying to find the L2 translation of a word in the first language). According to Laufer and Hulstijn (2001), the evaluation component "entails a comparison of a given word with other words, a specific meaning of a word with its other meanings, or comparing the word with other words in order to assess whether a word does or does not fit its context" (p. 14). Laufer and Hulstijn further state that evaluation is considered moderate when choosing between different words (as in a fill-in-the-blank task) or recognizing the differences between the several meanings that a word may mean in a specific situation. Evaluation is considered strong when decision has to be made about additional words that combine with new word in an original text. Each of these three factors can be absent or present when processing a word in a natural or artificially designed task. The combination of the above components and their degrees of prominence makes up involvement load expressed in terms of involvement index (moderate = 1, and strong = 2).

Laufer and Hulstijn's motivational-cognitive construct of task induced involvement is based on the framework of depth of processing, originally proposed by Craik and Lockhart (1972). According to depth of processing, the chance of some piece of new information being stored into long-term memory is not determined by the length of time that it is held in short-term memory, but rather by the shallowness or depth with which it is initially processed. In other words, Craik and Lockhart suggest that retention in long-term memory depends on how deep information is processed during learning.

Laufer and Hulstijn (2001) state that although Craik and Lockhart's depth of processing claims deeper processing leads to better memory performance, it says little about the actual mechanism of the processing, and it lacks an operational definition. For criticizing depth of processing theory, Laufer and Hulstijn set forth two questions: "(1) what exactly constitutes a level of processing? (2) How do we know that one level is deeper than the other?" (p.

5). In order to provide a more observable and measurable definition of depth of processing, they formulated the ILH.

The basic contention of the ILH is that retention of unfamiliar words is, generally, conditional upon the degree of involvement in the processing of those words. The concept of involvement can be submitted to empirical research by devising different tasks with various degrees of need, search, and evaluation. For example, tasks with different involvement indices can be presented to some groups of participants. Upon completion of the tasks, the results can be analyzed and compared to determine whether there is any relationship between the task involvement load and word retention.

Hulstijn and Laufer (2001) found empirical evidence for their hypothesis in a study conducted in an incidental learning environment. They investigated the effects of task-induced involvement via two parallel experiments. Learners in both experiments were randomly assigned to Task 1 (reading comprehension with marginal glosses), read a passage and answered multiple-choice comprehension questions that required knowledge of 10 target words. The target words were highlighted in the text and glossed in the margin. As a result, this condition called for moderate need (1), no search (0) and no evaluation (0). Overall, the involvement index was 1 (1+0+0). Participants assigned to Task 2 (reading comprehension plus fill-in) received the same reading passage and comprehension questions as in Task 1, but with the target words omitted and replaced with blank spaces. Learners were required to fill in the missing blanks using a list of words provided. The second task induced moderate need, no search, and moderate evaluation. Thus, the involvement index was 2 (1 + 0 + 1). Participants receiving Task 3 (composition writing) used the target words to write a letter. The meanings of the target words were glossed. With regard to the involvement load, this task induced moderate need, no search, and strong evaluation, as students had to incorporate the words into their compositions. So, its involvement index was 3 (1 + 0 +2). They found that the amount of retention was related to amount of task-induced involvement load as they predicted and was highest in the composition-writing task that induces highest level of involvement.

Kim (2008) also provided empirical evidence for the ILH in a carefully designed study consisting of two experiments. In experiment 1, three tasks with different involvement loads were given to ESL learners who were divided into

two language proficiency levels. Experiment 2 was intended to explore whether tasks with the same amount of involvement load produce the same results. The results of experiment 1 proved that higher level of involvement load leads to more effective initial learning and better retention of words. The findings of experiment 2 demonstrated that tasks with similar involvement load produced the same amount of immediate and delayed learning.

Jing & Jianbin (2009) studied the Involvement Load Hypothesis in incidental vocabulary acquisition in EFL listening. They gave three tasks to the subjects. Task A was listening comprehension questions with marginal glosses irrelevant to the questions (involvement index=0). Task B was listening comprehension questions with marginal glosses relevant to the questions (involvement index=1). Task C was listening comprehension questions with marginal glosses relevant to the questions and a composition writing (involvement index=3). They found that both in immediate and delayed tests, Task C with higher involvement load produced better vocabulary retention compared to Tasks B and A.

Folse (2006) studied the effects of different writing tasks on the learning of L2 words by university students whose proficiency levels ranged from lower intermediate to advanced level. Despite the overall support, he found that word learning to be more a function of repeated exposure than involvement. In other study, Martinez-Fernandez (2008) reported no support for ILH. She concluded tasks with higher degree of involvement load did not lead to deeper processing and higher vocabulary development. Thus, it seems necessary to conduct more studies before rushing to support the ILH claims.

Most of the empirical studies on the task-induced involvement load hypothesis have been concerned with reading-based vocabulary instruction on the experimentation of ILH, and very few studies on the task-induced involvement load have been conducted on integration of language skills. Therefore, the importance of the present research project lies in the fact that it is an attempt to fill the abovementioned gap in literature. The main goal of this study is to investigate the effect of performing two different tasks (i.e., jigsaw and information gap) with different degrees of involvement load on Iranian EFL learners' vocabulary recall and retention, and further to examine, which of the two tasks best promotes learners' recall and retention of the target wordsi

To achieve the purposes of the study, the following research questions were developed:

1) Do tasks with higher involvement load indices affect vocabulary recall of Iranian EFL learners?

2) Does the level of task-induced involvement affect the retention of new vocabularies of EFL learners when two tasks with different involvement loads are administered?

Method

Participants

The participants in this study were 64 female Iranian EFL learners from a private English language institute in Golpayegan, Iran. They were selected from a population of 80 learners based on their performance in Preliminary English Test (PET) which is a second level Cambridge ESOL exam for the intermediate level learners. The age of participants ranged from 19 to 28, and their first language was Persian. All the participants had studied English for at least six years at senior high schools and private language institutes. Then, they were randomly assigned into two experimental groups; one group as the jigsaw task group (Group A) and the other as the information gap task group (Group B). At the beginning, all participants were informed that attendance in ten treatment sessions over three-week period was obligatory.

Instrumentation

The materials used in this study were ten Passages from *The World of Words: Vocabulary for College Students* (Richek, 2011) course book, which is proper for intermediate level language learners as the author of the book agreed its suitability for intermediate (B1) learners. The length of each passage ranged from 500 to 700 words. Also, ten unfamiliar words from aforementioned book were presented on a sheet of paper (definition sheet) every session, including English definitions, their synonyms, their root meanings.

Moreover, the immediate and delayed posttests were administered to measure the participants' vocabulary recall and retention upon their completion of the two tasks with different involvement loads. Recall of unfamiliar words was measured immediately after the completion of vocabulary tasks. One month after the last treatment session, the learners received a delayed posttest,

based on the result of which, learners' long-term retention of the target words was assessed.

Target Words. The target words which were instructed and tested included nouns (28 words), verbs (28 words), and adjectives (44 words). The unfamiliarity of 100 target words was ascertained through these steps: First, forty-one students who were representative of the sample selected for the study were asked to choose unknown words on a list of 120 words. Based on their responses, 100 target words that were unfamiliar to all of them were selected for the study. Only one participant was excluded after the test because the learner knew some target words. Also, in the process of target word selection some other factors were considered, such as the lack of exposure to the target words outside the class, ease of supplying a synonym or definition in English, an appropriate translation in Persian, different parts of speech, as well as selection of almost lower frequent words.

Another rationale for the choice of unknown words was the etymological background of those words. Rivers (1981) states that knowledge of lexical roots assist in vocabulary retention because it helps students guess what a word means. In this study, the researchers were not interested in the high loads of information embedded in etymological studies, but the aim of the researchers was to facilitate and enhance vocabulary recall and retention via etymology.

Procedure

The first-named researcher of the present paper (henceforth the instructor) divided the participants into two groups, A and B, with different involvement loads containing 32 learners in each one. Two groups were provided with the same definition sheet and the same target words in ten sessions over three weeks. The target words were highlighted in bold type to be noticed, and they were introduced by the instructor within 20 minutes each session. Group A was instructed through the jigsaw task and Group B, information gap task. The groups were given the opportunity to get familiar with the tasks and practice the kind of activities they were supposed to receive. According to Ellis (2003) familiarity with the task is one of the factors that may promote learning more effectively. All instruction and assessment took place in the participants' regular class time. Group A attended the treatment sessions on odd days, while Group B on even days of the week.

Given the purpose of the study, each total group of 32 participants was divided into two sub-groups of 16, and then they got into groups of four in order to facilitate the completion of the task. The instructor monitored the groups during task completion. Time on task took about 80 minutes in Group A, and 60 minutes in Group B. An important issue to consider was the notion of time on task, since it is believed that "task effectiveness is a function of time spent on task" (Keating, 2008, p. 379). However, Hulstijn and Laufer (2001) consider time "as an inherent property of a task, not as a separate variable" (p. 549). Because this study intended to investigate retention effects of tasks, the researchers made no attempt to control for time on task.

Jigsaw Task with an Involvement Load of 3. At the beginning, the instructor explained the list of ten unknown words from the definition sheet. After teaching the list, the definition sheets were collected by the instructor in order to use dictionary by participants during the task. So, search component was present in this task. Then, the jigsaw task was carried out in the following steps: First, the instructor chose a passage from aforementioned course book. Second, the instructor divided the passage into four parts of nearly the same length, including unknown words for the participants. Then, the class was randomly divided into groups of four. For each group, the instructor appointed one student from each group as the leader, and assigned each student a special number. The participants were asked to remember their own numbers because the instructor arranged them by their numbers. Each group held one part of the passage to discuss it. The participants were encouraged to talk about the passage and consult with each other. The leaders of each group organized their own groups to discuss the assigned paragraphs, including summarizing the main ideas and comprehending all the sentences. The participants were guided by the instructor whenever they needed.

when each member of the groups fully understood the assigned paragraph, all the members became an expert in their groups. In this phase, the instructor asked all the participants to join to other jigsaw groups according to their numbers. For example, all number ones joined together and then all number twos and so on. Now each member in a jigsaw group had unique information, and the members of each jigsaw group had to teach each other their assigned paragraphs, respectively. They exchanged the information to form a totally

coherent passage. The parts of tasks, like different parts of puzzle, were completed through cooperation among all members.

Fill-in task was completed by participants as the immediate posttest. The involvement index of the task was 3(+1 need, + 1 search, + 1 evaluation). This task induced moderate need (1) because the need to learn the target words was imposed by the task. Search was present since the meaning of the target words were looked up in a dictionary. To control the variable, the participants used the same dictionary, *Oxford Advanced Learner's Dictionary* (2008). The task induced moderate evaluation since it required the participants to recognize differences between several meanings of a word, to choose the one that best matched the context, and to fill in the gaps with target words from the list.

Information Gap Task with an Involvement Load of 2. At first, the unknown words were instructed and then the participants were randomly divided into groups of four. The instructor designated one person for each group as a leader every session. Group leaders had important roles because they had certain information to share with others in order to perform the task. The leaders received the passage and read it in 10 minutes in order to explain the content to the other three members. In fact, the leader held the information about the passage that the other members of group did not have this information. Members of the group were only required to listen to the descriptions of the leader. They individually took notes and asked question whenever they needed. During the task and also in all test administrations, the instructor was present for clarifying the ambiguities for the participants.

The involvement index of the second task was 2(+1 need, 0 search, + 1 evaluation). It induced moderate need, imposed by the task, no search because participants were provided with glosses, and they did not have to look up the words in a dictionary, as well as moderate evaluation because the students had to evaluate whether a certain word fitted a given context. Both groups were tested on the target items once immediately after the instruction and the second time one month after the class. They took the same immediate and delayed posttest. The method of scoring was adopted in line with Hulstijn and Laufer's (2001) study in such a way that an incorrect or a blank answer to the item received no points or 0, and correct answer received 1.

Results

After the required data were collected, they were subjected to different quantitative analyses. The independent variable of the study was level of task-induced involvement and the dependent variable was the knowledge of the target words. In order to answer the research questions, the following statistical analyses were used. In order to examine how each group performed across the sessions on the immediate posttests, it was necessary to compare the means related to each session for each group by running repeated measures ANOVA. It should be noted that one of the assumptions of ANOVA is normality of the data, which was ensured by computing the skewness and kurtosis ratios (i.e., skewness and kurtosis values divided by their standard error). Since all these values were within minus/plus 1.96, it was concluded that the data were normally distributed.

Comparison of Immediate Posttests across the Sessions in Jigsaw Group

After checking the normality of the data as one of the assumptions of ANOVA, sphericity as another assumption was checked employing Mauchly's test, whose results indicated that this assumption was met (p > .05) (Table 1).

Table 1
Mauchly's Test of Sphericity[a]

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse -Geisser	Huynh -Feldt	Lower- bound
Session	.287	34.275	54	.985	.808	1.000	.100

In order to determine whether there is any significant difference among the sessions in Group A, pairwise comparison employing Bonferroni adjustment for multiple comparisons was run whose results demonstrated that the delayed posttest mean percentage was significantly lower than the immediate posttests mean percentage scores (p < .05). the results indicate, as the sessions passed by, the students in Group A recalled the words better on immediate posttests 3, 5, 9, 10. Although these results were significant, they need to be compared with those of the Group B to make necessary comparisons between the two groups.

Comparison of Immediate Posttests across the Sessions in Information Gap Group

After checking the normality of the data as one of the assumptions of ANOVA, sphericity as another assumption was checked employing Mauchly's test, whose results in Table 2 indicated that this assumption was not met (p < .05); therefore, sphericity was not assumed in ANOVA results, and the row labeled Greenhouse-Geisser was used in Table3.

Table 2
Mauchly's Test of Sphericity

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Session	.023	103.451	54	.000	.495	.600	.100

Table 3
Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Session	Sphericity Assumed	30234.455	10	3023.445	51.633	.000	.625
	Greenhouse-Geisser	30234.455	4.946	6112.795	51.633	.000	.625
	Huynh-Feldt	30234.455	5.998	5040.638	51.633	.000	.625
	Lower-bound	30234.455	1.000	30234.455	51.633	.000	.625
Error(Session)	Sphericity Assumed	18152.455	310	58.556			
	Greenhouse-Geisser	18152.455	153.329	118.389			
	Huynh-Feldt	18152.455	185.942	97.624			
	Lower-bound	18152.455	31.000	585.563			

Table 3 presents the main repeated measures ANOVA results. Evidently, the Greenhouse-Geisser row indicates that the ANOVA result is significant showing a significant difference somewhere among the sessions (p < .05).In order to see any significant difference among the sessions in Group B, pairwise comparison employing Bonferroni adjustment for multiple comparisons was

run. It demonstrated that the delayed posttest mean percentage is significantly lower than all the immediate posttest mean percentage scores (p < .05) exactly like what was found in Group A. However, there was no significant difference between the immediate posttests from session 1 to session 10 in terms of vocabulary recall scores (p > .05).

This finding can be significant since it is different from what happened in Group A. If remembered, as the sessions passed by, the students in the jigsaw group recalled the words better on the immediate posttests 3, 5, 9, and10. However, no difference was found among the sessions in Group B. This result can mean that as the students were given jigsaw tasks, they performed better on immediate recall on some of the subsequent sessions, but this did not happen at all in Group B.

Since the same words were taught to each group before immediate posttest, it was decided to compare the mean score on each immediate posttest in each group with that in the other group. This would allow seeing which task resulted in better immediate recall after each session of instruction. The choice of statistic for this comparison was multivariate analysis of variance (MANOVA) which would allow comparing the immediate posttest in each session across the two task groups.

Running MANOVA requires meeting several assumptions. The first of these is normality which was already checked in the previous sections. The second is ensuring that the dependent variables are subcategories of the same variable; in this study, all the dependent variables were related to general vocabulary knowledge divided into 10 parts. The third assumption is equality of covariance matrices, which was tested employing Box's test, whose results in Table 4indicate that it is met (p > .05).

Table 4
Box'sTest of Equality of Covariance Matrices[a]

Box's M	87.932
F	1.080
df1	66
df2	12256.740
Sig.	.308

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.
a. Design: Intercept + Group

The fourth assumption of MANOVA is equality of error variances tested by Levene's test, whose results indicate in Table 5. To resolve the violation of this assumption for the delayed posttest, a stricter p value such as .025 was considered in the main MANOVA results.

Table 5
Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Session 1	2.004	1	62	.162
Session 2	.362	1	62	.550
Session 3	.000	1	62	.988
Session 4	1.721	1	62	.194
Session 5	.667	1	62	.417
Session 6	.043	1	62	.837
Session 7	1.875	1	62	.176
Session 8	.980	1	62	.326
Session 9	1.243	1	62	.269
Session 10	.203	1	62	.654
Delayed posttest	6.788	1	62	.011

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Based on the descriptive already computed, the Group A with higher involvement load is of significantly higher mean scores on all the immediate posttests. Therefore, the first null hypothesis is rejected, that is to say, tasks with higher involvement load indices (i.e., jigsaw) affect vocabulary recall of Iranian EFL learners in the sense that jigsaw tasks result in better short term recall.

The second research question aimed at investigating whether the level of task-induced involvement affect the retention of new vocabulary of EFL learners when two tasks with different involvement loads are administered. The statistical findings showed that the two groups significantly declined on the delayed posttest. In this phase, it was necessary to first compute the average of all the mean percentage scores on the immediate posttests to come up with a representative score for all the immediate posttests. Then the average immediate posttest scores of all the sessions were compared against those of the delayed posttests to investigate long term retention, if any. In order to compare the average scores of all the immediate posttests and the mean of the delayed

posttest in each group, it was first necessary to compute the descriptive statistics and normality of the data.

The descriptive for the two groups indicate that the jigsaw scores on the average immediate posttest and delayed posttest are not deviant from normal since the skewness and kurtosis ratios are not beyond minus/plus 1.96 (Table 6); however, one of the skewness ratios calculated from the descriptive in Table 7 is beyond minus 1.96, hence violating of normality. Therefore, to compare the average immediate posttest scores and delayed posttest mean scores, paired samples t test was employed. The results indicated that in both groups the delayed posttest scores have declined significantly, showing no long-term retention of vocabulary in either group.

Table 6
Descriptive Statistics (Group A)

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
							Std. Error		Std. Error
Delayed posttest	32	38.00	80.00	56.5625	10.47559	.345	.414	.003	.809
sessions.average	32	82.00	94.00	87.7750	3.32643	.004	.414	-.875	.809
Valid N (listwise)	32								

Table 7
Descriptive Statistics (Group B)

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
							Std. Error		Std. Error
Delayed posttest	32	32.00	62.00	47.3750	6.33347	-.065	.414	.273	.809
sessions.average	32	67.20	86.80	79.2625	3.30178	-1.275	.414	5.755	.809
Valid N (listwise)	32								

Despite the significant decline on the delayed posttests scores, it was decided to investigate which task resulted in fewer declines and less forgetting. To do so, it was necessary to compare the two groups in terms of their delayed posttest means while their initial differences on the average immediate posttests scores were controlled. In so doing, the analysis of covariance (ANCOVA) was employed. Table 8 presents the main ANCOVA results, which shows that there is no significant difference between the two groups in long-term retention of the words on the delayed posttests ($p > .05$). In fact, the level of task-induced involvement shown in jigsaw and information gap tasks does not affect the retention of unknown words of EFL learners when two tasks with different involvement load are administered.

Table 8
Tests of Between-Subjects Effects

Source	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	2	1454.389	28.738	.000	.485
Intercept	1	539.116	10.653	.002	.149
Group * sessions.average (interaction)	1	74.173	1.477	.229	
sessions.average (covariate)	1	1558.216	30.789	.000	.335
Group (comparing delayed posttests)	1	80.578	1.592	.212	.025
Error	61	50.609			
Total	64				
Corrected Total	63				

a. R Squared = .485 (Adjusted R Squared = .468)

Discussion

The results provided support for the ILH in the first research question. Group A with higher involvement index was shown significantly higher mean score on all the immediate posttests. However, no difference was found among the sessions in Group B. This finding is in accordance with Hulstijn and Laufer (2001), Jing and Jianbin (2009), Keating (2008), and Kim (2008).

A probable reason for the better performance of Group A on immediate posttest might be the different degree of the search induced by jigsaw task. In this study search component was present (+search) in Group A but absent in Group B (- search). According to ILH, searching activity would increase the difficulty of the task which results in more cognitive process. Keating (2008)

states that recall of previously unknown words is better when words are looked up in a dictionary (+ search) as compared to when word meanings are glossed in the margin (no search). In addition, Nation (2001) asserts three important factors affecting L2 vocabulary development: noticing, retrieval, and the generation. He believes noticing can take place when a learner is looking up a word in dictionary.

Another reason for better recall in Group A was the task type. Participants in Group A had more opportunities to negotiate meaning during the task, and they worked better on integrating four language skills in task completion. From an involvement load perspective, some studies illustrate that words negotiated for meaning are retained better than no negotiated words (e.g., Ellis & He 1999; Fuente, 2002; Newton, 1995; Webb, 2005). Although information gap task required participants to use listening, speaking, reading, and writing to fill information gaps, there was less interaction in this task.

The second research question indicated significant declined in both groups' delayed posttest, and there was no significant difference between the two groups in long term retention of the words on the delayed posttests (p > .05). Similar to Folse's (2006) findings, this result was in contrast with the predictions of the ILH.

A plausible reason for significant decline in delayed posttest could be the large number of unknown words. The results of delayed posttest indicate that the large number of vocabulary strongly influence the retention effects. Also, the researchers also expected such a loss because the participants had no exposure to the target items in the interval between the two posttests that is, after a period of one month. Thus, it is not surprising that there would be a decline in retention for the task that initially showed the greatest gains. However, contrary to the predictions of the ILH, tasks with higher involvement load were as effective as task with lower involvement loads in long term vocabulary retention.

Some recent studies have suggested that the operational definitions of component constructs, (i.e., need, search, and evaluation) need to be modified (e.g., Baleghizadeh & Abbasi, 2013; Maftoon & Haratmeh, 2012; Martinez-Fernandez, 2008). As Kim (2008) states, "it is possible that all the three components might not be equal in contributing to vocabulary learning" (p. 313). For instance, the construct of search might be needed to be modified in

terms of operational definition. It is assigned (+1) value whenever present in a given task regardless the type (monolingual or bilingual) of the dictionary in which the unknown word is searched for. As Baleghizadeh and Abbasi (2013) state, the magnitude of involvement load might change with the type of the dictionary used by the learner and the involvement load generated as a function of working with a monolingual dictionary might be comparatively higher in magnitude than the one which is produced from a bilingual dictionary.

In terms of form-focused instruction, the results of this study supports the idea that word learning is better when vocabulary instruction includes a focus on form component. Ellis (2001) defines form-focus instruction as "any planned or incidental instructional activity that is intended to induce learners to pay attention to linguistic forms" (p. 2). According to Ellis (2003), explicit learning conditions and class room-based tasks designed to focus learners' attention on specific forms may be more effective for EFL classroom.

The findings of this study, from a pedagogical point of view, help EFL teachers to manipulate language components and to design tasks that enhance vocabulary learning. However, in addition to the involvement load, other factors such as task type should be considered in determining task effectiveness. With regard to the implications of the present study, EFL teachers can introduce the jigsaw and information gap tasks into their classroom, along with encouraging the learners to do dictionary-look up. This might not only lead to more productive use of the language, but also a correct example sentence can often be found in the dictionary as a good guide (Nasrollahy Shahry, 2010). The researchers particularly suggest jigsaw task because it is more beneficial, and when designed well, this task type is really challenging, engaging, and promotes a great deal of negotiation for meaning.

In order to validate the achieved results of the presented study, more research is needed. further studies can be conducted to compare the effectiveness of other types of task in vocabulary learning, among learners with different levels of English proficiency and with different or the same involvement loads. It would be possible to add more target words and more participants in a comprehensive study. In this study, the researchers measured the etymological feature of vocabulary learning. Other additional features, such as phonological, syntactic, semantic, and collocation should be taken into

account to examine the effect of task-induced involvement load on the retention of each one of these distinct features. Future research on the ILH should examine the long-term effect of the hypothesis by providing learners with multiple exposures to the target words. In fact, investigating the relationship between task types and the number of exposures may also shed light on the effectiveness of the hypothesis from a long-term perspective.

References

Baleghizadeh, S., &Abbasi, M. (2013). The effect of four different types of involvement indices on vocabulary learning and retention of EFL learners. *The Journal of Teaching Language Skills,5*(2), 1-26.

Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly*, *43*, 79-103.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework of memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-683.

Eckerth, J., &Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research, 16*, 227-252.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: OUP.

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

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

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.

Hedeen, T. (2003). The reverse jigsaw: A process of cooperative learning and discussion. *Teaching Sociology*, 325-332.

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*. Cambridge: CUP, 86-258.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty &M. H. Long (Eds.). *The hand book of second language research*, 349-381 Oxford: Blackwell.

Hulstijn, J. H., &Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539-558.

Jing, L., & Jianbin, H. (2009). An empirical study of the involvement load hypothesis in incidental vocabulary acquisition in EFL listening. *Polyglossia*, *16*, 1-11.

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

Keating, G. H. (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. (2008). The role of task induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *58*(2), 285-325.

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook, 5*, 223-250.

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.

Maftoon, p. & Haratmeh, M. S. (2012). Effect of Input and Output-oriented Tasks with Different Involvement Loads on Receptive Vocabulary knowledge of Iranian EFL Learners. *Iranian Journal of Research in English Language Teaching, 1* (1), 28-52.

Martinez-Fernandez, A. (2008). Revisiting the involvement load hypothesis: Awareness, type of task and type of item. *Selected Proceedings of the 2007 Second Language Research Forum*, 210-228.

Mengduo, Q., &Xiaoling, J. (2010). Jigsaw strategy as a cooperative learning technique: Focusing on the language learners. *Chinese Journal of Applied Linguistics*, *33*(4), 113-125.

Nagy, W, E., Hermn, P., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233-253.

Nasrollahy Shahry, M. N. (2010). *The effect of receptive and productive vocabulary learning through reading and writing sentences on vocabulary acquisition*. (Unpublished master's thesis). Shahid BeheshtiUniversity, Tehran, Iran.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: CUP.

Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research 11*(1), 159-177.

Nunan, D. (2004). *Task-based language teaching*. Cambridge: CUP

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

Pica, T., Sook Kang, H., &Sauro, Sh. (2006). Information gap tasks: Their multiple roles and contributions to interaction research methodology. *Studies in Second Language Acquisition*, *28*(2), 301-338.

Preliminary English Test (2006). Cambridge: University of Cambridge ESOL Examinations.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching*, Cambridge: CUP.

Richek, M. A. (2011). *The world of words: Vocabulary for college students,*(8th ed.). Illinois: Illinois University Press.

Rivers, W. M. (1981). *Teaching foreign-language skills.* Chicago, Illinois: University of Chicago Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12* (3), 329-363.

Shehadeh, A. (2005). Task-based language learning and teaching: Theories and applications. Inc. Edwards & J. Willis (Eds.). *Teachers exploring tasks in English language teaching* (pp. 13-30). Hampshire: Palgrave Macmillan.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing of word knowledge. *Studies in Second Language Acquisition, 27*,33-52.

Yaqubi, B., Rayati, R. A., & Gorgi, A. (2010). The involvement load hypothesis and vocabulary learning: The effect of task types and involvement index on L2 vocabulary acquisition. *The Journal of Teaching Language Skills*, *1*(1), 146-163.

Biodata

Atefeh Amini is a Master's graduate of TEFL (Applied Linguistics) at Islamic Azad University, Tehran Science and Research Branch, Department of Literature and Foreign language. Her research interests include language teaching methodology, second language acquisition, and Applied linguistics.

Parviz Maftoon is Associate Professor of teaching English at Azad University, Science and Research Branch, Tehran, Iran. He received his Ph.D. degree from New York University in 1978 in Teaching English to Speakers of Other Languages. His primary research interests concern EFL writing, second language acquisition, and syllabus design. He has published and edited a number of research articles and books. He is currently on the editorial board of several language journals in Iran.