

## The Effect of the Involvement Load Hypothesis on Improving Iranian EFL learners' Incidental Vocabulary Acquisition in Listening Comprehension Classes

NafisehAsadzadehMaleki

Malekan Branch, Islamic Azad University, Malekan, Iran

---

**Abstract:** 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. This study is aimed to apply the hypothesis to substantiate incidental vocabulary acquisition in EFL listening comprehension. To this end, 80 pre-intermediate EFL learners were assigned to one of the four groups: a control group that received no treatment and three experimental groups that completed one of the three vocabulary learning tasks that varied in the amount of the involvement; group 1, listening comprehension questions with marginal glosses irrelevant to the questions, group 2, listening comprehension questions with marginal glosses relevant to the questions, group 3, listening comprehension questions with marginal glosses relevant to the questions which were followed by writing sentences using those words. According to the immediate and delayed vocabulary retention post-tests' results, it turned out that there was a significant difference in retention effects among the three tasks, which proved the validity of the Involvement Load Hypothesis, confirming that tasks with higher involvement load lead to better retention effects. The findings of the current study have implications for language teachers, researchers, and material writers to encourage their students to use the involvement load hypothesis in order to enhance their vocabulary knowledge incidentally.

**Key words:** Listening Comprehension, Incidental Vocabulary Learning, Involvement Load, Depth of Processing, Retention.

---

### INTRODUCTION

When people think about communication, they usually focus on the ability to send and receive messages. Listening is important because the ability to understand what others say is essential to communicative interaction. Everyone who has ever learned a foreign language has probably experienced the frustrating feeling of not being able to communicate with native speakers of the language despite years of training in the target language. The main cause of this communication problem refers to the lack of the EFL learners' vocabulary knowledge. (Nagy, W.E., *et al.*, 1985) suggest that most words in both first and second languages are probably learned incidentally, through extensive reading and listening. Vocabulary development is very crucial, regardless of the purpose of the language learning. In fact, we need to use our vocabulary knowledge whenever we communicate. Vocabulary as a major component of language learning has been the object of numerous studies each of which has its own contribution to the field. (Laufer, B., 1997) states that vocabulary learning is at the heart of language learning and language use, Without vocabulary speakers cannot convey meaning and communicate with each other in a particular language (Laufer, B., 2003).

According to (Hulstijn, J.H., B. Laufer, 2001), words can be learned in the classroom through systematic teaching, or they can be picked up through listening or reading incidentally. In a comprehensive review of research on incidental learning, (Krashen, S., 1989) concluded that incidental vocabulary learning, or "acquisition", achieves better results than intentional vocabulary learning. A major flaw in this review lies in the assumption that "spelling and vocabulary are developed in second language as they are in the first language".

With regard to incidental vocabulary learning, recently, (Laufer, B., J.H. Hulstijn, 2001) tentatively proposed the notion of the Involvement Load Hypothesis. The authors hypothesize that retention of unfamiliar words depends upon the degree of involvement in dealing out of these words. Tasks with different involvement load will lead to different incidental acquisition (Laufer, B., J.H. Hulstijn, 2001). This hypothesis has been developed based on the depth of processing model, which was first proposed by (Craik, F.I.M., R.S. Lockhart, 1972). They recommended that retention in the long-term memory depends on how deeply information is processed during learning. Involvement is perceived as a motivational- cognitive construct, and retention of unfamiliar words is claimed to be conditional upon the amount of involvement while processing these words. (Laufer, B., J.H. Hulstijn, 2001) propose that the construct of involvement is composed of three components: need, search, and evaluation. Need is the motivational, non-cognitive component of involvement and refers to

whether knowledge of new words is required to complete a task. It is concerned with the need to achieve. While, according to (Schmidt, R., 1994), search and evaluation are the two cognitive dimensions of involvement, contingent upon allocating attention to form-meaning relationships. Search is the attempt to find the meaning of the unknown words or trying to find the second language word from expressing a concept by consulting a dictionary or another authority, e. g. a teacher. Evaluation entails 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. More to the point of determining these three components, (Laufer, B., J.H. Hulstijn, 2001) suggested three degrees of value for each component (none, moderate, and strong). According to (Laufer, B., J.H. Hulstijn, 2001), need can occur in two degrees of prominences: moderate and strong. Need is hypothesized to be moderate when it is imposed by an external agent, for example, when the teacher asks students to use a word in a sentence. Need is strong when it is intrinsically motivated and self-imposed by the learners, for instance, when learners decide to look up a word in a dictionary while writing a composition.

According to (Hulstijn, J.H., B. Laufer, 2001), although need can occur in two degrees of prominence (i.e. moderate or strong), search is not conceptualized as the relative degree of cognitive processing; instead it is either present (1) or absent (0). Search is present when learners must seek the meaning of the unknown words to complete a task (e. g. dictionary look-up task), and absent when no search effort is required (e. g. reading comprehension tasks accompanied by marginal glosses). Evaluation comprises of two potential degrees of cognitive processing: moderate and strong. Evaluation is moderate when learners must recognize differences between words provided in a given context (e. g. deciding which meaning of the target word best fits the context in which it is encountered) and strong when the task requires making decisions about new words and combining them with known words in original contexts (e. g. sentence and composition writing).

Empirical supports for the construct of task-induced involvement originates from the large body of literature on incidental word learning accomplished prior to its conception, and from recent researches designed to directly test the predictions of the Involvement Load Hypothesis. (Laufer, B., 2003) examined learning new words through writing sentences in contrast with learning new words through reading a text with glosses. She illustrated that using new words to write original sentences (strong evaluation) is more effective than reading new words glossed in sample sentences or reading a text for comprehension (no evaluation), regardless of whether new words are glossed (no search) or looked up (search). (Hill, M., B. Laufer, 2003) examined the effect of dictionary look-up tasks on the retention of unfamiliar words. These tasks were: task1, selecting the meaning of a new word in a text from several points, and task2, reading a text with glosses for comprehension. They demonstrated that task 1 was more effective than task 2 because in task 1 the evaluation index is 1 while in task 2 it is 0. (Newton, J., 1995) investigated the effect of interaction and negotiation in new vocabulary acquisition. He confirmed the efficacy of negotiation and interaction on vocabulary acquisition and found that new words negotiated in communicative tasks (strong need, search) are learned better than words that are not negotiated (no need or search). investigated the effects of task-induced involvement through two parallel experiments in advanced Dutch and Hebrew speaking learners of English. They presented three types of reading tasks: task 1 (reading comprehension with marginal glosses), task 2 (reading comprehension plus fill-in), task 3 (composition writing). They revealed that task 3 was more efficient than task 1 and task 2. However, task 2 was only superior to task 1 in the Hebrew – English experiment. (Shuhua, G., 2003) investigated the incidental vocabulary learning through reading comprehension skill with English majors. The author found that vocabulary acquisition is affected by presenting different types of tasks and different students' vocabulary size as well. (Joe, A., 1995) observed the effects of different degrees of generation on vocabulary learning. He showed that words with high degree of generation are better retained than words with a low degree of generation, or no generation at all. According to the involvement load hypothesis, what this means is that words which were retained better underwent a higher involvement load than words which were retained less well.

Accordingly, a key issue in second language vocabulary learning was whether or not learning would be improved if information about an individual lexical item were to be presented simultaneously via increasing the amount of the learners' involvement load. There are some studies about the efficacy of using various vocabulary acquisition ways, but it seems that there are little studies focusing on the effect of the involvement load hypothesis on vocabulary learning especially through listening comprehension in Iranian EFL context. Therefore, motivated by previous studies on vocabulary learning and the involvement load hypothesis and existing gaps in this literature, this study was designed to broaden our understanding of the application of the involvement load hypothesis in a foreign language acquisition environment. It investigated the effect of the application of this hypothesis on learning incidental vocabularies through listening comprehension skill.

#### ***Research Questions:***

The present study was an attempt to answer the following research questions:

1. Do tasks with different degrees of involvement load have a different effect on incidental vocabulary acquisition of Iranian EFL learners' in immediate tests in listening comprehension classes?

2. Do tasks with different degrees of involvement load have a different effect on incidental vocabulary acquisition of Iranian EFL learners' in delayed tests in listening comprehension classes?
3. Are there any significant differences among Iranian EFL learners' incidental vocabulary acquisition in performing tasks with different degrees of involvement load hypothesis?

***Method:***

***Participants:***

In this study, 80 Iranian university students from Islamic Azad University, Maragheh Branch were participated. They were chosen from among 120 students according to their English language proficiency test scores. Thus, four parallel classes of freshman, pre – intermediate, English major students, with 20 in each, were selected for the research. The subjects were supposed to take a Preliminary English Test (PET) to check whether the general language proficiency level of the participants in the classes was relatively at the same level or not. Then each class was assigned to a specific kind of treatment: the control group that received no treatment and three experimental groups that received three different kinds of listening comprehension tasks.

***Instrumentation:***

***Experimental Passages:***

In this study, twelve listening comprehension passages were chosen from Longman Complete Course for the TOEFL Test, preparation for the computer and paper tests (2001). The lengths of the passages were relatively equivalent, in which the length of each passage ranged from 200 to 300 words. The selected passages were transcribed by the researcher and were modified to accommodate the target words.

***Experimental Tasks:***

In the current study, three tasks with different involvement load were designed for each class for investigation. Students in different groups were required to do different tasks. These three tasks are as follows:

Task A: listening comprehension questions with marginal glosses irrelevant to the questions.

Task B: listening comprehension questions with marginal glosses relevant to the questions.

Task C: listening comprehension questions with marginal glosses relevant to the questions followed by sentence writing with the same given target words.

***The Target Words:***

In this research, words were one of the most important components. The researcher concentrated on four most common word types: nouns, verbs, adjectives and adverbs. These words were considered new to most of the students, because in the pre - experiment vocabulary test these words were tested with the students. The researcher presented the target words in a table. The table consisted of four columns: target words were included in column one, then, in column two, the subjects were required to write YES if they had seen the target words before. In the third column, the subjects were supposed to write NO if they had not seen the target words before. Finally, in column four, the subjects were supposed to write the meaning of the words if they knew. The results showed that most of the students could not recognize them, since approximately 32.06 % of words were familiar to the subjects.

***Pre – Test:***

In this study, the researcher administered a pre – test. The subjects were asked to complete a multiple choice test. In this test, 40 target words were selected from the listening comprehension passages and were presented in their original context taken from Longman Complete Course for the TOEFL Test, preparation for the computer and paper-based tests (2001). For each word the subjects were asked to choose the correct answer out of four given choices. Of the four choices, one was the correct meaning, and the other three were distracters. In this test, the correct answer scored 1 and a score 0 was given for an incorrect or not attempted answer.

***Immediate Post-Test:***

To answer the first research question which was designed to investigate the effect of the involvement load hypothesis on incidental vocabulary learning in an immediate post-test, a multiple choice test consisting of 40 target words was designed as a measure of learners' developing knowledge of particular words through applying different task type. Again in this test, 40 target words were presented in their original context taken from Longman Complete Course for the TOEFL Test, preparation for the computer and paper tests (2001). In the same way as the pre – test, the subjects were required to answer the multiple choice questions and choose the correct meaning of the target word among four given choices. Of the four choices, one was the correct answer, and the other three were incorrect answers.

**Delayed Post-Test:**

In order to measure the retention of the target words, the researcher gave a delayed post-test to the same subjects. The same target words of the immediate test were presented in this study. However, in the delayed test, the researcher presented the same target words in different contexts. The sentences that included the target words were selected from the Oxford Advanced Learner's Dictionary (2003). For each word, the subjects were asked to choose one correct meaning out of four given choices.

**Design:**

The design to carry out this research was quasi – experimental, with different treatments for the experimental groups. The experiment was conducted within sixteen 90 minutes class sessions. Immediately after the treatment sessions, the subjects were asked to take an immediate test to assess the effect of each task on vocabulary learning. The test was a multiple choice recognition test consisting of forty items from the listening passages. Two weeks later, in order to assess the retention of the target words, the subjects were required to take a delayed test which was as the same as the immediate test, but the target words were presented in different context.

**Procedure:**

In this study, 80 homogenous Iranian EFL learners participated. The subjects were supposed to take a Preliminary English Test (PET) to assess their level of language proficiency. The results showed that most of the students were at the same level of proficiency, therefore, each class was randomly assigned to one of the four groups, so there were three experimental groups and one control group. First, the subjects met with the researcher. They were given a brief introduction to the program, its objectives, and its methods. Then, the researcher provided twelve listening comprehension passages for the students. During the treatment sessions, three different tasks with different involvement load were planned for each of the classes in experimental groups as follows:

Task A: listening comprehension questions with marginal glosses irrelevant to the questions. In this kind of task, the students were supposed to listen to the passage, and then they were asked to answer the multiple-choice questions. The students were able to answer the questions without understanding the unknown words. According to [6], in this case, task A induced no need, because students did not need the target words to answer the questions; no search, because all the target words have been glossed for them, and no evaluation, because there was not any chance for them to compare and decide the meaning of the word in comparison with different words or different senses of the same word, so the involvement index was 0.

Task B: listening comprehension questions with marginal glosses relevant to the questions. After listening to the passage, the students were asked to complete multiple-choices, which can only be answered with the understanding of the unknown words. In task B, there was a moderate need, because the need to learn the target words was imposed by the task; no search, because as in task A, all the target words have been glossed for them; no evaluation, because they were not required to compare and decide on the meaning of the word in comparison with different words or different senses of the same word, so the involvement index was 1.

Task C: listening comprehension questions with marginal glosses relevant to the questions followed by sentence writing with the same target words. The first part of the task was the same with task B, the students were asked to complete multiple-choice questions. For the second part of the task after answering the listening comprehension questions, the subjects were required to write sentences with the given target words. In this case, still, there was no search, because all the target words have been glossed for them; there was a moderate need, because the need to learn the target words was imposed by the task; there was a strong evaluation, because students had to make a decision as to how additional words could be combined with the new words in a sentence or even a text. Therefore, the involvement index was 3

Class	Class I	Class II	Class III
Task	Task A	Task B	Task C

According to the table, subjects in class 1 were required to perform task A, the subjects in class 2 were asked to perform task B, and finally, the subjects in class 3 were asked to perform task C. Then, after completion of the tasks, the researcher collected all the work sheets and the subjects were given a sheet of a test paper that involved the target words with their original context picked out of the passages. In this test, which is known as the immediate test, the subjects were required to choose the correct meaning of the bolded target words in each sentence among four distracters. After the immediate post-test, the data sheets were collected by the researcher and the items were scored 0, if the students provided nothing or incorrect answers to the questions, and they were scored 1, if the correct and exact answers were chosen by the subjects. Two weeks later, the subjects were asked to take a delayed test to assess their retention of the target words. The researcher presented the same target words of the immediate test in different context and the students were required to

recall the meaning of those words and choose the correct meaning of the words from among four distracters. Finally, the collected data were analyzed to assess the retention of the target words by the researcher.

**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.

**Results:**

In order to answer the aforementioned research questions, 120 Iranian university students participated in this study. In order to check their level of proficiency, the subjects were required to take a Preliminary English Test (PET). According to their English language proficiency scores, 80 subjects were chosen to fulfill this study. Afterwards these subjects were divided into four groups. Their proficiency scores were calculated via utilizing a one-way ANOVA to determine the homogeneity of the groups. Table 4.4 demonstrates the descriptive data for the proficiency scores of the groups.

**Table 1.** Descriptive Data for the Proficiency Tests, PET, of the Groups

	N	Mean	Std. Deviation	Std. Error
Class 1	30	12.90	3.52	0.64
Class2	30	12.96	4.03	0.73
Class 3	30	13.00	3.93	0.71
Class 4	30	13.10	4.07	0.74
Total	120	12.99	3.84	0.35

**Table 2.** ANOVA Test of within Groups' Effects for Proficiency Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.625	3	0.208	0.014	0.998
Within Groups	1762.367	116	15.193		
Total	1762.992	119			

According to the descriptive data in Table 4.4, the mean score of the groups are approximately at the same level, and the results of the ANOVA test in Table 4.5 show that the Sig. ratio of the proficiency scores is (f (3,116) = 0.014, p = 0.998), therefore it demonstrates that there is not a significant deference among the four groups.

The first research question of this study aimed to examine the effect of the involvement load hypothesis through presenting different tasks with different degrees of the involvement on incidental vocabulary acquisition in immediate tests in listening comprehension classes as follows:

1. Do tasks with different degree of involvement load have a different effect on incidental vocabulary acquisition of Iranian EFL learners' in immediate tests in listening comprehension classes?

To answer the first research question, the scores of each type of task for each class were submitted to a one-way ANOVA. The results of this analysis are shown in Table 4.6.

**Table 3.** Descriptive Data for the Immediate Test of the Groups

	N	Mean	Std. Deviation	Std. Error
Control	20	28.95	2.91	0.65
Task A	20	30.10	2.80	0.62
Task B	20	31.15	3.01	0.67
Task C	20	32.40	2.16	0.48
Total	80	30.65	2.98	0.33

The results in Table 4.6 illustrate that the mean score for Task C is (32.40), which is higher than the mean scores for the other groups. However, the mean score for Task B (31.15) is higher than the mean score of Task A and control group. As well, the mean score for Task A (30.10) is higher than the mean score for the control group (28.95). Therefore, the means displayed in Table 4.4 show that the subjects in Task C group have performed better than the subjects of the other groups. In addition, the subjects in Task B group had performed better than the others, and Task A group was better than the control group. Table 4.7 shows the ANOVA results of the immediate test scores.

**Table 4.**ANOVA Tests of within Groups' Effects for Immediate Test Scores

	Sum of Squares	df	Mean Square	f	Sig.
Between Groups	130.10	3	43.36	5.76	0.001
Within Groups	572.10	76	7.52		
Total	702.20	79			

As it is clear in Table 4.7, the result of the ANOVA indicates that there is a statistically significant difference among the performances of the groups on the immediate test ( $f(3,76) = 5.76, p = 0.001$ ), since P value is lower than 0.05. Therefore, the results confirmed that the involvement load hypothesis had a significant effect on improving Iranian EFL learners' incidental vocabulary acquisition in immediate tests in listening comprehension classes. Also, it can be concluded that from among different tasks with different degrees of involvement, Task C indicated more significant effect than the other tasks on promoting EFL learners' vocabulary acquisition.

The second research question investigated the effect of the involvement load hypothesis through giving the learners an idea about different tasks with different degrees of the involvement on enhancing EFL learners' incidental vocabulary acquisition in delayed test in listening class. As a result, the second research question of this study was:

2. Do tasks with different degree of involvement load have a different effect on incidental vocabulary acquisition of Iranian EFL learners' in delayed tests in listening comprehension classes?

The obtained data of the delayed tests of the groups were analyzed through utilizing a one-way ANOVA. Table 4.8 demonstrates the descriptive statistical data for the performances of the groups on the delayed post-test.

**Table 5.**Descriptive Data for the Delayed Post-Test of the Groups

	N	Mean	Std. Deviation	Std. Error
Control	20	29.35	3.08	0.68
Task A	20	30.65	2.56	0.57
Task B	20	32.00	2.47	0.55
Task C	20	33.70	1.59	0.35
Total	80	31.42	2.92	0.32

As it is demonstrated in Table 4.8, the mean score for the subjects in Task C group is (33.70) and it is higher than the other groups. Accordingly, through comparing the mean scores of the groups it is clear that the Task C group outperformed all other groups on the delayed post-test. Moreover, the subjects in Task B group (with the mean score of (32.00)) have performed better than the subjects in Task A group (30.65) and the subjects in control group (29.35). According to Table 4.8, it was clarified that there is a difference between the performances of the groups on the delayed post-test. Table 4.9 indicates the ANOVA results of the delayed post-test scores.

**Table 6.**ANOVA Test of within Groups' Effects for Delayed Test Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	208.25	3	69.41	11.24	0.000
Within Groups	469.30	76	6.17		
Total	677.55	79			

The results in Table 4.9 yielded a significant effect for the involvement load hypothesis ( $f(3,76) = 11.24, p = 0.000$ ) on improving EFL learners' incidental vocabulary acquisition on the delayed post-test. From among the different tasks, Task C designated more significant effect on increasing incidental vocabulary acquisition of Iranian EFL learners in a listening comprehension classes.

The exploration of the effect of the involvement load hypothesis on EFL learners' incidental vocabulary acquisition in listening class was the concern of the third question:

3. Are there any significant differences among Iranian EFL learners' incidental vocabulary acquisition in performing tasks with different degrees of involvement load?

The difference between the pre-test scores and immediate post-test scores, and also the differences between the pre-test scores and delayed post-test scores were analyzed through Post Hoc Tests.

According to Table 4.12, there is a significant difference between the mean score of the control group and the mean score of the Task C. Therefore, it can be claimed that the subjects in Task C group outperformed the subjects in the control group, Task B group. Table 4.10 also indicates that there is a significant difference between the scores of the subjects in Task A group and Task C group. Hence, it enlightened the extra effect of the Task C than the Task A. What's more, the post-hoc analysis of mean differences among the pre-test and the immediate-test revealed that there was not a significant differences between the scores of Task B group and other groups, i.e. control group, Task A group, and Task C group. Regarding the acquired results from Table

4.10, it can be concluded that Task C was more effective than the other tasks in learning and retention of new target words in immediate test.

**Table 7:** Post Hoc Analysis of Mean Differences between Pre-test and Immediate-test

(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Control	Task A	-0.600	0.828
	Task B	-1.600	0.112
	Task C	-3.000*	0.000
Task A	Control	0.600	0.828
	Task B	-1.000	0.489
	Task C	-2.400*	0.006
Task B	Control	1.600	0.112
	Task A	1.000	0.489
	Task C	-1.400	0.200
Task C	Control	3.000*	0.000
	Task A	2.400*	0.006
	Task B	1.400	0.200

Multiple Comparisons, Dependent Variable: DiffImmediate

**Table 8:** Post Hoc Analysis of Mean Differences between Pre-test and Delayed-test

(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Control	Task A	-0.750	0.880
	Task B	-2.050	0.188
	Task C	-3.900*	0.001
Task A	Control	0.750	0.880
	Task B	-1.300	0.576
	Task C	-3.150*	0.014
Task B	Control	2.050	0.188
	Task A	1.300	0.576
	Task C	-1.850	0.269
Task C	Control	3.900*	0.001
	Task A	3.150*	0.014
	Task B	1.850	0.269

Multiple Comparisons, Dependent Variable: DiiDel

According to Table 4.15, the results of the post-hoc analysis of the mean differences among four groups in both pre-test and immediate-test are, also, the same as the results of this analysis in both pre-test and delayed-test. This table shows that there is a significant difference between the scores of the Task C group and the control group. As well, the mean difference between scores of the Task C and the Task A is significant. Therefore, it can be declared that task C has an effect on learning and retention of new vocabularies incidentally in listening classes.

**Discussion:**

This study was an attempt to provide an empirical evidence for the involvement load hypothesis in second language incidental vocabulary acquisition in listening comprehension classes. The first and the second research questions of this study addressed the effectiveness of the involvement load hypothesis on immediate and delayed tests through examining different types of tasks with different degrees of involvement on improving EFL learners' incidental vocabulary learning and retention. The third question compared the effect of different tasks with different degrees of involvement load. 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.

In this study, learners in the experimental groups performed better than the learners in the control group. EFL learners who completed Task C outperformed the learners who completed Tasks A and B. Through post-hoc analysis, it was confirmed that there was a significant difference between the scores of the pre-test and immediate-test among the Task C group, Task A group, and the control group. This result, also, was proved in the case of the differences between pre-test and the delayed-test. In this study, however, the difference between Task C and B results was not significant. Consequently, it is reasonable to authenticate the effectiveness of the hypothesis on incidental vocabulary learning and retention in an Iranian EFL context.

The results of the study partially confirmed the results of Hebrew-English experiment in (Hulstijn, J., 2001), who investigated the effect of the involvement load hypothesis on vocabulary learning through reading comprehension skill. They found that Task 2 and 3 resulted in higher retention scores compared to Task 1. In their experiment, there was a significant difference between Task 3 and 2. However, in the current study, the difference between Task B and Task C was far from significant in the post-hoc test (p = 0.20 in the immediate test, and p = 0.26 in the delayed test). Hulstijn and Laufer's Task 3 required learners to use target words to write

an original composition while in the current study subjects used them to write their own sentences with the given target words.

(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.

The results of this study are also consistent with those of (Rott, S., 2005). She paid close attention to the involvement load hypothesis. Her research is intended as an investigation of how words are processed to strengthen lexical form-meaning connections of four target words during a reading exercise. Two different conditions were determined in her study: single-translation gloss condition and multiple-choice gloss condition. In her study, the multiple-choice gloss facilitated more need, search, and evaluation than the single translation gloss, and the former also appeared to bring about stronger form-meaning connection and also more result memory trace than the latter. As the results of the current study, in Rott's study, the tasks that needed higher degree of involvement were resulted in more retention of the new words.

The outcomes of this study were also in line with (Hulstijn, J.H., 1992) study related to the effects of the involvement load hypothesis on vocabulary learning and retention. He showed that when meanings of words had to be inferred they were retained better than words with given meaning. In his study, regarding the involvement load, in synonym-condition the value index for evaluation is 0 while in multiple-choice condition it is 1. Therefore, the involvement index in multiple-choice condition is higher than the former. Consequently, the results of this study confirms the results of the current study, determining that tasks with higher degree of involvement lead to better learning and retention of new words.

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.

Similarly, the results of this study proved the findings of (Hulstijn, J.H., P. Trompetter, 1998) research. They examined two separate tasks to learn new words: task A which refers to learning new words through writing composition, and task B that refers to learning new vocabularies via reading comprehension text. They concluded that task A was more efficient than task B because it requires strong evaluation.

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.

As far as the researcher knows among the studies concerned the concept of the involvement load hypothesis, the study with the findings which were incompatible with the current study was that of (Martinez-Fernandez, A., 2008) study. He examined the involvement load hypothesis by focusing on awareness, type of task and type of items. He identified four different types of tasks, and revealed that tasks with higher degree of involvement load did not lead to deeper processing, defined as high awareness, and did not lead to higher vocabulary development.

### **Conclusion:**

There has been a revival of interest in issues of vocabulary learning in a second language in the last decade or so and a significant amount of empirical research has been carried out on the subject. Vocabulary learning is often used with strategies such as word lists or paired associations in which new words are presented with their translations. As mentioned in the preceding chapters, the main focus of this study was on improving EFL learners' incidental vocabulary acquisition through the involvement load hypothesis. Previous studies have examined the effects of the involvement load hypothesis on second language vocabulary learning through reading comprehension. These studies have sustained the effectiveness of this hypothesis in facilitating second language vocabulary learning. Regarding the review of the related literature in this area, this study has considered the efficiency of the above-mentioned hypothesis in the domain of incidental second language vocabulary acquisition through listening comprehension skill.

The results of this study revealed that the task with higher involvement load, in this research, namely, task of multiple-choice questions with marginal glosses relevant to the questions followed by sentence writing tasks secured the best retention effects, task of multiple-choice questions with marginal glosses relevant to the task produced better retention outcomes than task of multiple-choice questions with marginal glosses irrelevant to the questions, which produced the lowest retention.

The outcomes of the current study generally corroborated (Laufer, B., J.H. Hulstijn, 2001) findings. Their experiments showed that the task of composition with incorporated target words produced best retention results, and the task of reading comprehension plus filling in target words produced better results than task of reading comprehension with marginal glossing for target words.

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.

(Laufer, B., J.H. Hulstijn, 2001) states that "processing new lexical information more elaborately (e.g. by paying attention to the word's pronunciation, orthography, meaning) lead to higher retention than by processing new lexical information less elaborately (e.g. by paying attention to only one or two of these dimensions)" (p. 270).

To conclude, the involvement load hypothesis can be applied in learning and retention of new target vocabularies, it is also applicable in incidental vocabulary acquisition. The involvement load hypothesis should be familiar to language teaching researchers and professionals far more than it is now, since it covers a variety of factors, and it is essential to have solid vocabulary learning theories to use in classroom situations. Indeed, this hypothesis needs to be investigated in studies and classrooms so as to lessen the burden of vocabulary learning. Consequently, the involvement load hypothesis posits that incidental tasks that induce higher involvement are conducive to the type of processing that is considered crucial for vocabulary retention.

Accordingly, the involvement load hypothesis will be a first step in stimulating theoretical and empirical work and will be followed by other proposals of greater scope and depth. Greater depth could be reflected in more precise definitions of the involvement components and a more through theoretical link between them and theories of information processing and affective components of cognition.

In conclusion, the involvement load hypothesis is applicable to the incidental vocabulary acquisition. The study demonstrated that in listening practice, incidental acquisition does occur and task with higher involvement load does produce better retention.

## REFERENCES

- Craik, F.I.M., R.S. Lockhart, 1972. Levels of processing: A framework for memory research, *Journal of Verbal learning and Verbal Behavior*, 11: 671-684.
- Hill, M., B. Laufer, 2003. Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition, *International Journal of Applied Linguistics*, 41: 87-106.
- Hulstijn, J., 2001. Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In P. Robiunson (Eds.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.
- Hulstijn, J.H., 1992. Retention of inferred and given word meanings: Experiments in incidental word learning. In P. J. L. Arnaud, H. Béjoint (Eds.) *Vocabulary and Applied Linguistics* (pp. 113-125). London: Macmillan.
- Hulstijn, J.H., B. Laufer, 2001. Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51(3): 539-558.
- Hulstijn, J.H., P. Trompetter, 1998. Incidental learning of second language vocabulary in computer-assisted reading and writing tasks. In D Albrechtsen, B Henriksen, I. M Mees, EPoulsen (Eds.), *Perspectives on foreign and second language pedagogy: Essays presented to Kirsten Haastrup on the occasion of her sixtieth birthday* (pp. 191-200). Odense, Denmark: Odense University Press.
- Joe, A., 1995. The value of retelling activities for vocabulary learning. *Guidelines*, 17: 1-8.
- Keating, G.D., 2008. Task effectiveness and word learning in a second language: the Involvement Load Hypothesis on trial. Retrieved on November 24, 2009 from <http://ltr.sagepub.com>
- Krashen, S., 1989. "We acquire vocabulary and spelling by reading: additional evidence for the Input Hypothesis". *Modern Language Journal*, 73: 440-464.
- Laufer, B., 1997. What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N Schmitt, M McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.
- Laufer, B., 2003. Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review*, 59: 567-587.

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

Martinez-Fernandez, A., 2008. Revisiting the Involvement Load Hypothesis: awareness, type of task and type of item. *Second Language Research Forum*, 20: 210-228.

Nagy, W.E., P.A. Herman, R.C. Anderson, 1985. Learning words from context. *Reading Research Quarterly*, 20: 233-253.

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

Rott, S., 2005. Processing glosses; a qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, 17: 95-124.

Schmidt, R., 1994. Implicit learning and the cognitive unconscious: of artificial grammars and SLA. In N. C. Ellis (Eds.) *Implicit and explicit learning of languages*. London: Academic Press, pp: 165-210.

Shiping, D., Y. Chensong, 2004. The effects of multiple annotations on incidental vocabulary acquisition. *Foreign Language Teaching and Research (bimonthly)*, 36: 213-218.

Shuhua, G., 2003. Empirical study on incidental vocabulary acquisition of English majors. *Foreign Language Teaching and Research (bimonthly)*, 35: 282-286.