# Vocabulary Acquisition and Task Effectiveness in *Involvement Load Hypothesis*: A case in Iran

Hassan Soleimani

Department of Applied Linguistics, Payame Noor University, Tehran, Iran

E-mail: arshia.soleimani@gmail.com

Mahboubeh Rahmanian (Corresponding author)

Payame Noor University, Tehran, Iran

E-mail: rahmanian74@yahoo.com

## Abstract

Involvement load hypothesis as a cognitive construct states that tasks with higher involvements yield better results in vocabulary retention. This comparison group designed study examined the immediate and delayed effects of tasks with different involvements in involvement load hypothesis (Laufer & Hulstijn, 2001). Applying a version of Nelson Proficiency Test as a homogenizing exclusion criterion, 33 low proficiency Iranian EFL learners were randomly assigned to three experimental groups: blank-filling, sentence making, and reading comprehension. The results of *ANOVA* and *Kruskal-Wallis* tests supported task-induced involvement in immediate posttest since the sentence making task ($M$=5.72) yielded better results in comparison with the other two blank-filling ($M$=5.45) and reading comprehension ($M$=3.18) tasks. Nevertheless, sentence making and blank-filling tasks of which the involvements were somehow similar did not yield significant superiority to each other. It is inferred that tasks with nearer involvements yield somehow similar results in vocabulary acquisition.

**Keywords:** task-induced involvement, immediate effect, delayed effect, vocabulary retention, evaluation

## 1. Introduction

Attention is a concept of great concern to many theories in cognitive psychology and language learning issues including the *noticing hypothesis* (Schmidt, 1990, 1994), limited processing ability (VanPatten, 1990), and "pushed output" (Swain, 1985). Attention which is the necessary and sufficient perquisite for long-term retention often relates to models of memory (Schmidt, 1994). Murdock's (1967) modal memory holds that attention needs to be devoted to a stimulus in order to be processed through different stages. However, in 1972, Craik and Lockhart mention some defects in this model that in effect produced their two-store model. Their *depth of processing hypothesis* holds that the three levels of orthographical, acoustic, and semantic should be dealt with for a deeper processing of a stimulus and consequently its better retention (Baddeley, 1999). This hypothesis which deals with the internal processing stages of learning a stimulus in mind holds that this depth of processing can have the outcome of more durable and firmer traces for learning new items. They bring us two boxes: sensory memory holding information which has gone through threshold analyses; Short-term memory (STM) holding information which has gone through deeper analyses. Along the same lines, Laufer and Hulstijn (2001) added some other processing levels to Craik and Lockhart's (1972) claim. Understandably their *involvement load hypothesis*, which considers three components of *need*, *search*, and *evaluation* for a task, holds that the more the involvement load indexes are, the more elaborate the acquisition of vocabularies will be. Although Laufer and Hulstijn's (2001) *task-induced involvement* can be regarded as a comprehensive and original construct, its predictions may not be so much relied on and more complementary studies (e.g., Hulstijn & Laufer, 2001; Xu, 2009; Walsh, 2009) are needed. This study is determined to see if supportive or contradictory evidence can be found for involvement load hypothesis in terms of its immediate and delayed effects.

## 2. Literature Review

One of the important issues which many researchers have pointed out is the stream of consciousness and the wave of attention. Titchener (1910, as cited in Martindale, 1991) notes that *consciousness* can be "arranged in to focus and margin, foreground and background, center and periphery" (p. 266). Martindale (1991) considers his definition and states that *attention* is the focus, the foreground, and the center. Schmidt (1990) in his noticing hypothesis believes, "that noticing is necessary for SLA, and that understanding is facilitative but not required" (p. 725). He mentions the work by Logan, Taylor, and Etherton (1996) who state that when a learner pays enough attention to a stimulus, it will be encoded for further processing in the learner's mind. Craik and Lockhart's (1972) depth of processing hypothesis enumerates three factors that can affect deeper processing and as a result more effective learning, that is attention, the

time which is devoted to the processing of each stimulus, and accommodating to the previous schemata. Ryan (n.d.) points to the hidden layer of Swain's (1995) noticing the gap hypothesis as attention and states that this can make learners deal better with language competency like grammar. Hulstijn and Laufer (2001) claim that "[I]t is generally agreed that retention of new information depends on the amount and the quality of attention that individuals pay to various aspects of words" (p. 541). On the whole the concepts of cognition, levels of processing, and attention contributed Laufer and Hulstijn (2001) to set up their involvement load hypothesis.

Involvement load hypothesis which was suggested by Laufer and Hulstijn (2001) holds that the effort that an individual devotes to a task mentally or its involvement load is considered to be the determinant factor in learning. It regards some levels for a stimulus to be processed and in effect retain longer, that is it starts "with shallow sensory analysis, and proceeding to deeper, more complex, abstract, semantic analysis" (Solso, 1988, p. 133). This incidental learning theory is developed in line with the focused instruction of vocabulary.

Arguably, this motivational and cognitive innovation of task-induced involvement includes the three elements of need, search, and evaluation. Tasks with higher involvement indexes result in more effective vocabulary learning. Understandably tasks can have different values of absence, "moderate", and "strong" for each of these elements. Need or the motivational element of involvement concerns with the motivation of completing a task. It can be moderate when it is task-imposed or strong when it is self-imposed. Search and evaluation or the cognitive elements of involvement concern with information processing or attending to new items. Search is the attempt a learner devotes to find out the meaning of an unfamiliar word. It can be present when the meaning of words needs to be searched or absent when they are provided in the marginal glosses. Evaluation, on the other hand, refers to considering context and putting the best unfamiliar word in (moderate evaluation) or composing a new sentence by it (strong evaluation). Evaluation which entails coming to an appropriate solution for the meaning of an unfamiliar word is "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). The combining of all these components with their load involvement indexes can specify the most appropriate task for learning a word item. Laufer and Hulstijn (2001) give much of their attention to tasks with higher load involvement indexes since in accordance with involvement load hypothesis these tasks can lead to higher vocabulary acquisition in comparison with tasks with lower load involvement indexes.

In order to determine the load involvement index of a task we can add up the values of its components. Hulstijn and Laufer (2001) suggested the following procedure: "absence of a factor is marked as 0, a moderate presence of a factor as 1, and strong presence as 2" (p. 544). Therefore, a task can have zero to five involvement indexes. In order to better reveal these elements in classroom environment, some examples can be resorted to. A task in which students are required to (moderate need) compose a sentence (strong evaluation) with provided glosses (no search) has an involvement load of three. In fact by adding up the degrees of each component, we come in to the task's involvement index. In another task, we have a student answer some multiple-choice questions after reading a passage which has some glosses in its margin. Here, the involvement index is one since moderate need is the only present factor. Task-induced involvement claims that the first example leads to longer retention since its involvement index is higher. This operationalization can be done by manipulating different tasks in classrooms to help the students learn better. Laufer and Hulstijn (2001) complete their innovative theory by stating that the only determinant factor in the retention of new vocabulary items is the involvement index of the tasks and no other factor like their proficiency level or task type (i.e., input or output) is beneficial. In other words, they believe that no task type has priority over another.

The supportive evidence for the innovative construct of task-induced involvement has a tradition as long as incidental vocabulary learning. However, the studies which were designed to directly test involvement load hypothesis are a few. To mention some, the experiments by Hulstijn and Laufer (2001), Kim (2011), Jing and Jianbin (2009), Folse (2006) Keating's (2008) can be named that brought supportive and contradictory evidence for involvement load hypothesis. Hulstijn and Laufer (2001) developed the first empirical investigation concerning their innovative notion of involvement load hypothesis. Their study was designed in order to unravel the effect of task-induced involvement in a short and long run for retention of 10 unknown vocabulary items. Advanced Dutch-Hebrew EFL participants were considered to reveal the impact of three learning tasks with different load involvement indexes: marginal-glossed task with an involvement of one, fill-in-blank task with an involvement of two, and composition-writing task with an involvement of three. Immediately after the administration of tasks, an immediate posttest containing the ten target items was given to the learners in order to observe the initial influence of tasks in terms of their vocabulary retention. One to two weeks later, the similar posttest was administered for testing the delayed influence of tasks. The results of the two posttests supported Laufer and Hulstijn's (2001) task-induced involvement in the three tasks of marginal-glossed, fill-in-blank, and composition-writing. However, the marginal-glossed and fill-in-blank tasks were not significantly better than each other in spite of their different load involvements.

Kim (2011) examined the involvement load hypothesis considering different task types and proficiency levels. He tried to observe the impact of three tasks with different involvement loads in two different levels of proficiency. Reading, gap-fill, and composition tasks were randomly assigned to the participants in each proficiency group. Two immediate and delayed posttests were administered in order to examine if any short-term or long-term vocabulary retention has been resulted from the experiment. The composition group with an involvement index of three brought about significantly better results in comparison with the reading group with an involvement index of one and gap-fill group with an involvement index of two. Nevertheless, the results for the gap-fill and reading groups in delayed posttest

revealed the superiority of the gap-fill group over the reading group. In order to explain the results, the predictions of Laufer and Hulstijn (2001) can be modified in some delicate ways. In other words, involvement load hypothesis came true in all its dimensions only for the delayed posttest and not the immediate one. Furthermore, Kim (2011) had another experiment in which he studied two tasks with equal load involvement indexes. He compared writing composition task and writing sentence task with three involvement indexes in order to see if these tasks can lead to similar retention of vocabularies. They caused equal vocabulary retention in terms of their initial and long-term recall in both immediate and delayed posttests. Thus, another supportive evidence was gained for the task-induced involvement of Laufer and Hulstijn (2001).

Jing and Jianbin (2009) examined Laufer and Hulstijn's (2001) involvement load hypothesis to reveal if its predictions can come true for the listening comprehension tasks too. The participants were required to answer some comprehension questions in the three tasks. Task A in which some marginal glosses of the target word items were prepared asked the learners to answer some questions which answering them did not require knowing the word items (0+0+0=0). Task B, on the other hand, asked the learners to answer some questions which answering them required knowing the word items (1+0+0=1). Task C which was the same as Task B asked the learners to write a short article with the target words too (1+0+2=3). The immediate and delayed posttests of the study helped us to investigate this hypothesis more profoundly.

Folse (2006), on the other hand, showed contradictory evidences for the involvement load hypothesis. He examined the influence of different writing tasks on vocabulary retention and found that tasks of moderate and strong evaluation yield similar results in terms of vocabulary acquisition. Understandably, writing new sentence task which had strong evaluation component brought about the same results as gap-fill task which had moderate evaluation component.

Low-proficiency Spanish learners were the participants of Keating's (2008) study who were randomly assigned to three tasks with different involvements. The results of her immediate and delayed tests revealed the higher retention of the second and the third tasks with two and three involvements respectively in comparison with the first task with one involvement index. Nevertheless, the second and the third tasks were not so much better than each other. Keating's (2008) study indicated that the predictions of involvement load hypothesis can be generalized to the low proficiency learners. Keating (2008) added "Tasks that induce greater involvements (i.e., tasks with higher degrees of need, search, and evaluation) generally lead to greater gains in short-term and, in some cases, long-term word retention" (p. 368). As can be inferred, the delayed effect of tasks with higher involvement indexes is more ambiguous in comparison with its immediate effect. We are going to observe the immediate and long-term retention of tasks with different involvements.

*2.1 Research Hypothesis*

The following null hypothesis was formulated in order to unravel the effect of passage of time on participants' vocabulary recall in tasks with different involvement loads.

Null Hypothesis: There is no statistically significant difference among low proficiency EFL learners across three blank-filling, sentence making, and reading comprehension tasks with different load involvement indexes in immediate and delayed posttests.

## 3. Method

This experimental research used comparison group design because of the random assignment of the participants to the three groups.

*3.1 Participants*

In order to have three homogeneous samples, 66 male and female intermediate students between 19-25 years old with a mean age of 22 from two English institutes in Isfahan, Iran, were selected. Since each institute let us have three classes containing 11 students, two institutes were considered. Afterwards, Nelson Proficiency Test was administered in order to classify them into high and low proficiency groups. Then, using this test as an exclusion criterion in this research the low proficiency group was considered in order to have participants of the same level of proficiency. In parallel with the purpose of our study the three tasks, which had been prepared previously, were randomly assigned to the students in the low proficiency group.

3.1.1 Homogeneity Test Results

The normal distribution of the students' scores in Nelson Proficiency Test ($p$ = .096, p > .05) let us use the parametric statistic test one-way *ANOVA*. Based on its Levene test of homogeneity of variances, the homogeneity of the three groups was concluded, $F$ (2, 30) = .099, $p$ = .906 (2-tailed) at the α=.05.

*3. 2 Materials*

Three tasks based on a reading comprehension passage in line with ten target word items composed our materials in order to reveal task-induce involvement.

3.2.1 Tasks

Based on Laufer and Hulstijn's (2001) task-induced involvement, three tasks with different involvement indexes were prepared. The reading comprehension passage, which was the main part of each, was selected according to the suitability of its content for our target population. Our participants' teachers and some intermediate students of the mentioned institutes were provided with three reading passages. Being within the participants' readability level and general knowledge were among the factors which we regarded in the selection procedure of the reading passages. The reading text which Walsh (2009) had used in a study with similar purposes was selected.

Task A which had an involvement index of two was the blanked-out reading passage considering the target vocabulary items. Moreover, ten L2 marginal glossing words were provided for the learners in the first group to make the meaning of the words clear. This blank-filling task was administered to make learners put the most appropriate words in the spaces and in addition answer its reading questions. Its involvement index of two was the sum of a moderate need, no search, and moderate evaluation (1+0+1=2).

Walsh (2009) passage with its ten L1 glossing target words and five reading comprehension questions formed Task B. Unlike the previous task, the learners had to make a sentence with each target word. The sentence making a task which induced a moderate need, no search, and strong evaluation on the part of the participants had three involvement indexes. That is, the highest involvement index with regard to the three tasks (1+0+2=3).

The reading comprehension task which contained the passage with the ten L1 glossing target words and five multiple-choice questions made the learners only answer the questions using the glossing words. According to Laufer and Hulstijn (2001), Task C had an involvement index of one for its moderate need, no search, and no evaluation (1+0+0=1).

### 3.2.2 Target Vocabulary Items

The 326-word *Child Labor* reading passage of Walsh (2009) with a readability level of 7.73 also was given to the participants' teachers and those intermediate students to select the words which they knew as unknown for our target sample. Finally, all of the words which they selected as unknown compromised the 36-word pretest which brought us a reliability coefficient of .62. The following ten words were ultimately put as the most suitable target items: two verbs, six nouns, one adjective, and one adverb (*plantation*, *fair*, *demonstrations*, *crops*, *sweatshop*, *fiber*, *partly*, *blame*, *march*, and *shrimp*).

### 3.2.3 Data Collection Procedure

Thirty three low proficiency students were selected based on Nelson Proficiency Test. The three tasks were administered to the participants based on a random assignment. As a consequence, the learners' level of proficiency and age were of the same range to bring us three homogeneous groups. Testing the learners unexpectedly and introducing tasks as a reading exercise were executed to let us come up with the incidental vocabulary acquisition requirements (Laufer & Hulstijn, 2001). The researchers clarified how to complete the tasks for each group of the students. Furthermore, after the treatment, in accordance with Walsh (2009), we did not allow the participants to exchange whatever they had learned from the tasks in order to be much safer about the values of each component of tasks. This in effect could decrease the threat to internal validity. The learners completed the tasks in fifteen, twenty, and ten minutes for the first, second, and third tasks respectively. The immediate and delayed posttests were given to them two days and two weeks after the tasks completion, respectively. The scoring procedure for the posttests, which contained ten target vocabularies with different arrangements, differed from previous testing of vocabularies. As a matter of fact, zero to one point was devoted to each incorrect to correct equivalents of the target words, respectively. Furthermore, half a point was considered for related answers.

### 3.3 Data Analysis

In order to examine the normal distribution of the data, graphical and statistical tests were run. The K-S normality index revealed a normal distribution for the immediate posttest (p>.05, *p* = .080), and thus the parametric statistic test one-way *ANOVA* was utilized. However, the delayed posttest could not confirm the normal distribution (p>.05, *p* = .031), and in effect nonparametric *Kruskal-Wallis Test* was run.

## 4. Result

*Testing the Null Hypothesis*

RH0: There is no statistically significant difference among low proficiency EFL learners across three blank-filling, sentence making, and reading comprehension tasks with different load involvement indexes in immediate and delayed posttests.

Descriptive statistics for the immediate posttest which are displayed in Table 1 made it clear that the sentence making group received the highest mean scores in comparison with the blank-filling and reading comprehension groups.

Table 1. Descriptive statistics on retention scores of the immediate posttest

|                       | N  | M    | SD   |
|-----------------------|----|------|------|
| Blank-filling         | 11 | 5.45 |      |
| Sentence Making       | 11 | 5.72 | 1.48 |
| Reading Comprehension | 11 | 3.18 | .87  |
| Total                 | 33 | 4.78 | 1.59 |

In order to reveal if the scores of the immediate posttest were significant, one-way *ANOVA* was run (Table 2). The *ANOVA* results for the research hypothesis showed a significant difference among the means, F (2, 30) = 16.72; $p$ = .000 < .05. The sentence making task resulted in a significantly better retention.

Table 2. ANOVA for comparison of the means of the immediate posttest

|  | SS | Df | MS | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 42.97 | 2 | 21.48 | 16.72 | .000 |
| Within Groups | 38.54 | 30 | 1.28 |  |  |
| Total | 81.51 | 32 |  |  |  |

The *Tukey* post hoc test (Table 3) indicated that the mean score of the sentence making group was significantly different from the mean score of the reading comprehension group ($p$ =.000). However, the differences between the mean scores of the sentence making and blank-filling groups were not significantly different ($p$ =.840). The means of the blank-filling and reading comprehension groups were significantly different ($p$ =.000).

Table 3. Tukey post hoc for the retention scores of the immediate posttest

| (I) group | (J) group | Mean Difference (I-J) | SE | Sig. |
|---|---|---|---|---|
| Blank-filling | Sentence Making | -.27 | .48 | .840 |
|  | Reading Comprehension | 2.27$^*$ | .48 | .000 |
| Sentence Making | Blank-filling | .27 | .48 | .840 |
|  | Reading Comprehension | 2.54$^*$ | .48 | .000 |
| Reading Comprehension | Blank-filling | -2.27$^*$ | .48 | .000 |
|  | Sentence Making | -2.54$^*$ | .48 | .000 |

* $p$ < .05 level.

In order to test the delayed effect of the three groups of blank-filling, sentence making, and reading comprehension *Kruskal-Wallis Test* was utilized. Tables 4 and 5 clarify a statistically significant difference, $\chi^2$ (2, $n$= 33) = 18.87, $p$ = .000 in the vocabulary retention of the learners.

Table 4. Descriptive statistics on retention scores of the delayed posttest in terms of ranks

|  | Group | N | Mean Rank |
|---|---|---|---|
| delayed posttest | Blank-filling | 11 | 20.23 |
|  | Sentence Making | 11 | 23.68 |
|  | Reading Comprehension | 11 | 7.09 |
|  | Total | 33 |  |

Table 5. Kruskal Wallis Test for comparison of the means of the delayed posttest

|  | delayed posttest |
|---|---|
| Chi-square | 18.87 |
| Df | 2 |
| Asymp. Sig. | .000 |

*Note*. Grouping Variable: tasks.

Median scores of the learners in delayed posttest were the same in blank-filling and sentence making groups (*Md* = 5). However, they were higher than those of the reading comprehension group (*Md* = 2) which can be seen in Table 6.

Table 6.  Median of the three groups on the retention scores of the delayed posttest

| Group | N | Mdn |
|---|---|---|
| Balnk-filling | 11 | 5.00 |
| Sentence Making | 11 | 5.00 |
| Reading Comprehension | 11 | 2.00 |
| Total | 33 | 4.00 |

Running *Mann-Whitney U Test* which helped us find the location of the differences needed a Bonferonni adjustment to control for Type 1 errors. Consequently, the alpha level of .05  was divided by 3 (the number of pairs to be compared) calculating a stricter level of .017. The difference was between the sentence making and reading comprehension groups ($U$ = 1.00, $z$ = -3.980, $p$ = .000 < .017, $d$ = .69) and also between the blank-filling and reading comprehension groups ($U$ = 11.00, $z$ = -3.358, $p$ = .001 < .017, $d$ = .58) with large and medium effect sizes, respectively (Table 7 and Table 8).

Table 7. Mean ranks of the sentence making and reading comprehension groups of the delayed posttest

|  | delayed posttest |
|---|---|
| Mann-Whitney U | 1.00 |
| Wilcoxon W | 67.00 |
| Z | -3.980 |
| Asymp. Sig. (2-tailed) | .000 |

*Note*. Grouping Variable: tasks.

Table 8. Mean ranks of the blank-filling and reading comprehension groups of the delayed posttest

|  | delayed posttest |
|---|---|
| Mann-Whitney U | 11.00 |
| Wilcoxon W | 77.00 |
| Z | -3.358 |
| Asymp. Sig. (2-tailed) | .001 |

*Note*. Grouping Variable: tasks.

## 5. Discussion and Conclusion

The results of the experiment considering the mean scores of the learners in immediate and delayed posttests were somehow the same. The hypothesis, in fact, examined how tasks with different involvement indexes resulted in target vocabulary acquisition of EFL learners with the same level of proficiency. This study which reflects Laufer and Hulstijn's (2001) task-induced involvement can not support it exclusively. Understandably, the results of the research hypothesis, which explored the effect of involvement load hypothesis on the immediate and delayed posttests, indicated that Laufer and Hulstijn's (2001) claim can come true but the superiority of sentence making task is not so much significant compared to blank-filling task with lower but nearer involvement index.

In the parallel experiments by Hulstijn and Laufer (2001) and Kim (2008), it was found that the composition writing task yielded higher learning of target words than the other two tasks of glossing and gap-fill ones. However, gap-fill task did not yield significantly better learning than glossing task. We can declare that our results were somehow in line with their experiment concerning the better retention of tasks with higher involvement load.

In order to explain these results, we can observe the evaluation construct in more details. Since the sentence making and blank-filling tasks had their only difference in the evaluation component of task-induced involvement, its moderate and strong forms can come in to the scene. Sentence making task with strong evaluation and blank-filling task with moderate evaluation yielded approximately the same results. This is not only contradictory to the predictions of Laufer and Hulstijn's (2001) who believe in the equal contribution of components of task-induced involvement (i.e., need, search, and evaluation), but also to the predictions of Kim (2008) who mentions the unequal contribution of its components.  In a nutshell, Kim (2008) claims that tasks with strong evaluation involve learners more in processing

vocabulary items than those with moderate ones. It goes without saying that our findings were similar to those of Folse (2006) who found that tasks with strong and moderate evaluation were the same in terms of learning new vocabularies.

Table 9. Results of the previous and recent studies concerning the component of involvement load hypothesis

| Involvement Load Hypothesis | | Laufer and Hulstijn (2001) | Folse (2006) | Kim (2008) | Our study (2013) |
|---|---|---|---|---|---|
| Need | moderate | 1 | 1 | 1 | 1 |
| | strong | 2 | 2 | 2 | 2 |
| Search | moderate | 1 | 1 | 1 | 1 |
| | strong | 2 | 2 | 2 | 2 |
| Evaluation | moderate | 1 | 2 | 2 | 2 |
| | strong | 2 | 2 | 4 | 2 |

Although our study supported the predictions of task-induced involvement in both initial and long-term processing of target words, it was unraveled that tasks with nearer involvements are not so much superior to each other. Furthermore, it can be declared that this hypothesis can work similar for both immediate and delayed effects of vocabulary retention. This idea is contrary to Keating's (2008) who does not believe so much in the usefulness of involvement load hypothesis for longer retention.

*5.1 Implications and Limitations of the Study*

It is implied from the present study that tasks with higher involvements are better for pedagogical purposes. That is, tasks which induce learners to process the target words more profoundly have a good chance to be remembered. However, tasks with high but nearer load involvement indexes can not be predicted to conform to the principles of involvement load hypothesis. That is, similar results can be obtained using tasks with high but nearer involvement indexes. As can be observed, Laufer and Hulstijn's involvement load hypothesis is so baffling that applying its predictions in all tasks may not be true. In fact the involvement index of tasks need be taken in to account in order to come to a more appropriate conclusion about learning unfamiliar words in context.

Although our study looked at task-induced involvement in more details, it is exposed to some limitations. First, this study has only investigated low proficiency learners who are more in need of learning basic vocabularies. Second, we assessed the learners' receptive learning of words and their productive learning remained intact. Third, as Folse (2006) claims the effect of tasks with higher involvements can be neutralized or conversed if more than usual time is devoted to each task. As a result, it can be declared that since the time we devoted to each task might not be within the realm of the standards, the obtained results can not be considered highly reliable. Fourth, teachers' attitude, students' psychology (Lee, 2003), and type of teachers' reinforcement (Hulstijn & Laufer, 2001) can be regarded as the other factors which can affect vocabulary learning. As these factors were not considered, we can not be confident if our results were merely because of the load involvement indexes of the tasks. Thus, more confirmatory researches need to be conducted in order to reveal the real effect of involvement load indexes of tasks on vocabulary retention.

**Acknowledgements**

**References**

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

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

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

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

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

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

Lee, C. (2003). *The effect of involvement load on English vocabulary acquisition in an incidental learning situation* (Unpublished master's thesis). National Tsinghua University, Hsinchu, Taiwan.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory,and Cognition*, *22*, 620-638.

Martindale, C. (1991). *Cognitive psychology: A neural-network approach*. Pacific Grove, CA: Brooks Cole.

Murdock, B. B. JR. (1967). Recent developments in short-term memory. *British Journal of Psychology*, *58*, 421-433.

Ryan, J. (n.d.). A review of the role of output in second language acquisition with anecdotal examples from a Japanese learner's experience. Retrieved April 11, 2013 from leo.aichi-u.ac.jp/~goken/bulletin/pdfs/NO17/RyanJ.pdf.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129-158.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. In J. H. Hulstijn & R. Schmidt (Eds.), Consciousness in second language learning. *AILA Review,11*, 11-26.

Solso, R. L. (1988). *Cognitive psychology* (2$^{nd}$ ed.). Boston, Mass.; London: Allyn and Bacon.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidhofer (Eds.), *Principles and practice in the study of language* (pp. 125-144). Oxford University Press.

Titchener, E. B. (1910). *A text book of psychology*. New York: Macmillan.

VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, *12* (3), 287–301.

Walsh, M. I. (2009). *The involvement load hypothesis applied to high school learners in Japan: Measuring the effects of evaluation* (Unpublished master's thesis). Birmingham University, Edgbaston, United Kingdom.

Xu, J. (2009). An experimental study on the effects of different reading tasks on L2 vocabulary acquisition. *English Language Teaching*, *2*, 69-76.