

*The Journal of Teaching Language Skills (JTLS)*  
*Vol. 2, No. 1, Spring 2010, Ser. 59/4*  
(Previously *Journal of Social Sciences & Humanities*)

**The Involvement Load Hypothesis and Vocabulary Learning: The  
Effect of Task Types and Involvement Index on  
L2 Vocabulary Acquisition**

**Dr. B. Yaqubi\***  
Assistant Professor  
University of Mazandaran, Mazandaran  
email: yaqubi@umz.ac.ir

**Dr. R.A. Rayati**  
Assistant Professor  
University of Mazandaran, Mazandaran  
email: r.rayati@umz.ac.ir

**N. Allemzade Gorgi**  
M. A. student  
University of Mazandaran, I. R. of Iran

**Abstract**

This study builds on Laufer and Hulstijn's (2001) motivational-cognitive construct of task-induced involvement in learning vocabulary and addresses itself to its strong claim that the depth of processing is the overriding factor in learning words. The paper first re-examines the effect of processing load and then of task type on the initial learning and retention of words. To do so, 60 EFL learners from two branches of an English institute were selected. The participants were then randomly assigned to three groups: The first group completed an input-oriented task with an involvement index of three; the second group also completed the same type of task but with an involvement index of two, and the third group completed an output-oriented task with the same involvement load as that of the first group. The comparison of the performance of the groups in the immediate and delayed posttests reveals that contrary to the prediction of the involvement load

---

Received: 9/5/2009

Accepted: 4/24/2010

\* Corresponding author

**hypothesis, Task 2 with an involvement index of two was superior to Task 1, which had a higher index. Besides, the participants who had completed the *output oriented task* (Task 3) outperformed those that did the input-oriented task (Task 1), despite their index equivalency. The study suggests that the operationalization of the levels of processing, especially *evaluation*, needs reconsideration.**

**Keywords:** 1. Involvement Load Hypothesis 2. Task-Induced involvement 3. Involvement Index 4. Task 5. Retention 6. Need 7. Search 8. Evaluation

### 1. Introduction

For both language teachers and learners vocabulary is obviously a top priority (Schmitt, 2008). Having an extensive vocabulary is believed to help learners “to outperform their competence” (Nunan, 1999: 103), that is, a sizable vocabulary having been learned by L2 learners is contended to enable them to partly handle unpredictable communicative situations. For language learners, big words (i.e., content words) are indisputably the very elements that make it possible for them to figure out meaning (VanPatten, Williams, and Overstreet, 2004). Despite the importance of vocabulary in L2 learning, there has been conspicuously less theoretical headway in this domain than, for instance, grammar teaching (Laufer and Huljistin, 2001; Nation, 2001). However, in recent years due to the Noticing hypotheses (Schmidt, 1990, 1994, 2001), limited processing ability (VanPatten, 1990), and ‘pushed’ output (Swain, 1985), which to a large extent laid the foundation for Focus on Form, vocabulary learning researchers have also begun to advocate focused instruction of vocabulary. Involvement load hypothesis (Laufer and Huljistin, 2001), although not at first formulated in the context of form-focused instruction, claims that in incidental learning situations the retention of forms (words) depends on the manipulation of the cognitive and motivational variables within tasks. This claim can arguably constitute a technique in form-focused instruction which in effect is very different from the default position taken in vocabulary learning and teaching.

Despite the ambitious claim made by the proponents of involvement load hypothesis (ILH), there are many questions yet to be addressed. The present study is primarily concerned with the hypothesis itself and the link that might exist between the hypothesis and task types, a relationship that may shed more light on the claimed predictive power of involvement load hypothesis. More specifically, the study considers whether task type with different distributions of involvement load indexes, in this case, output-oriented tasks versus input-oriented tasks with equal involvement load, has similar effects on word retention.

## 2. Literature Review: Depth of Processing

Laufer and Hulstijn's (2001) motivational-cognitive construct of task-induced involvement is based on the framework of depth of processing, originally proposed by Craik and Lockhart (1972) as an alternative to multistory model (Atkinson and Shiffrin, 1968). Unlike Atkinson-and-Shiffrin's model in which stimuli are claimed to get processed at three relatively independent stages, the involvement load hypothesis argues that stimuli are processed at several *levels*, "starting with shallow sensory analysis, and proceeding to deeper, more complex, abstract, semantic analysis" (Solso, 1988: 133 ). And deeper levels of processing contribute to more elaborate, more durable, more meaningful, and stronger memory traces. In other words, retention is claimed to be less correlated with the length of time spent on a piece of information held in primary memory than by the depth to which the information is initially processed. In the same vein, Craik and Lockhart (1972) suggest that "trace persistence is a function of depth of analysis, with deeper levels of analysis associated with more elaborate, longer lasting, and stronger traces" (p. 675).

Primary stages are concerned with the superficial analysis of physical and sensory features like lines, pitch or loudness. Later stages are more concerned with matching the input against stored abstraction from past learning. That is, later stages are concerned with pattern recognition and the extraction of meaning. For example, seeing a car in the street, we may only come to know of its presence but totally ignore it. This is the shallowest

possible level of processing which may well not contribute to our remembering the car at a later time. At the deeper level we may also notice what make it is (Peugeot, Golf, etc.), the labeling strategy, which in turn is a deeper level of processing, and comes to our assistance in remembering it later (Olson and Hergenbahn, 2009). At the deepest level, we afford the car a new meaning, noticing its details are surprisingly like the one we used to have. The deeper we process the details of a stimulus, the more likely it is that we remember it later.

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 (Laufer and Hulstijn, 2001). Eysenck (2004) sees the hypothesis more as a descriptive account than an explanatory framework. Summing up the criticisms about processing depth (Nelson, 1977; Baddely, 1978; Eysenck, 1978), Hulstijn and Laufer (2001: 541) set forth two epistemological questions: "(1) What exactly constitutes a level of processing? (2) How do we know that one level is deeper than the other?" Despite the criticisms cited against the framework which had Craik and Lockhart's (1975) acknowledgement, Hulstijn and Laufer (2001) suggest, "Yet, it 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" (541).

### **2.1 Involvement load hypothesis**

As Schmidt (2008) contends, deeper engagement with new vocabulary as induced by tasks clearly increases the chances of their being learned. In their search for an operationalizable definition of the depth of processing theory (Craik and Lockhart, 1972), Laufer and Hulstijn proposed the motivational-cognitive construct of involvement which comprises three basic involvement components: *need*, *search*, and *evaluation*. The *need* component refers to the motivational, noncognitive dimension of involvement. Two degrees of prominence for *need*, moderate and strong, are differentiated in terms of the extrinsic-intrinsic distinction. For example, a *need* is moderate when it is imposed externally (e.g., the need to identify a

word in a sentence as has been asked by the teacher). A strong *need* is one that is self-imposed (e.g., a need induced as a consequence of the learners' decision, for example, to look up a word when writing a composition). *Search* and *evaluation*, as opposed to *need*, are the two cognitive dimensions of involvement, which are conditional upon the allocation of time to the form-meaning relationship (Schmidt, as cited in Hulstijn and Laufer, 2001). *Search* refers to "the attempt to find the meaning of an unknown L2 word from expressing a concept (e.g., trying to find the L2 translation of an L1 word) by consulting a dictionary or another authority (e.g., teacher)" (Hulstijn and Laufer, 2001: 14). *Evaluation* on the other hand, refers to the assessment of an appropriate meaning or use of a given word within the suitable context. Simply the component of *evaluation* implies "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 (i.e., a form-meaning pair) does or does not fit its context" (14). For instance, in a reading task in which an L2 word looked up has multiple meanings, the choice of the most appropriate meaning should be achieved by comparing all the meanings against the context. The component of *evaluation* has two degrees of prominence: When *evaluation* requires the use of a new word within a given sentence it is moderate, but when the learners are required to produce an original sentence, *evaluation* is strong because learners should judge how to combine words and produce a sentence.

According to Hulstijn and Laufer (2001), all of the three components may not be present simultaneously during a reading-based task. The combination of these factors with their degrees of prominence makes up the involvement load. Based on involvement load hypothesis, different tasks seem to impose varying demands on learners but we cannot compare different tasks and judge their involvement load based solely on intuition. As a result, researchers attempted to operationalize the abstract concept of involvement load into a measurable concept of "task-induced involvement". According to Hulstijn and Laufer (2001), "The basic contention of involvement load is that retention of unfamiliar words is generally

conditional upon the degree of involvement in processing these words” (545).

In their attempt to operationalize depth of processing, Hustijn and Laufer proposed “involvement index” in which the absence of a factor is marked 0, a moderate presence of a factor equals 1, and a strong presence of a factor equals 2. For example, consider two different tasks. In the first task, the learner is asked to write original sentences with some new words whose meanings are provided by the teacher. In this case the *need* is moderate (imposed by the teacher), there is no *search* (meanings are provided), and strong *evaluation* is required in that the learner has to use the new words by his/her own generation. In terms of involvement load the task has an involvement index of 3 (1+0+2). In the second task, the learner is required to read a text (with glosses of the new words) and to answer comprehension questions. The task induces a moderate *need*, but neither *search* nor *evaluation*. The involvement index of this task is 1 (1+0+0). To be more clear the first task is thought (according the construct of task-induced involvement) to induce a greater involvement load than the second task.

The concept of involvement can be submitted to empirical investigation by devising incidental-learning tasks with various degrees of *need*, *search* and *evaluation*. For example, tasks with different involvement indexes can be presented to some groups of participants. After they have finished the tasks the results can be analyzed and compared to see if there is any relationship between the task involvement load and word retention.

## 2.2 Empirical investigations

Laufer and Hulstijn (2001) bring it to our attention that a great deal of support for involvement load hypothesis predates its formulation by studies not run to test their hypothesis. Understandably, research studies having a direct bearing on the hypothesis are few and far between due to its recent formulation. Hulstijn and Laufer (2001) conducted two parallel experiments in which their advanced Dutch- and Hebrew participants (adult English learners) were formed into six intact groups. Retention of ten unfamiliar words in incidental learning setting was investigated across three

tasks types (Task 1 included reading comprehension with marginal glosses, Task 2 comprehension plus filling in target words, and Task 3 composition-writing with target words). The tasks had different involvement loads, i.e. various combinations of need, search and evaluation. The result indicated that Task 3 was more involving and led to better retention than Task 1 and Task 2, thus providing strong support for ILH.

Using eight nonsense words, Keating (2008) also used three tasks (Task 1 consisted of a reading passage with marginal glosses; Task 2 reading comprehension plus fill-in; Task 3 writing original sentences using target words) with different involvement loads to assess the predictive nature of the ILH, i.e., whether the hypothesis can be extended to low-proficiency learners. In Task 1 the low-proficiency participants had to read a passage with five true/false comprehension questions. To correctly answer the questions participants had to attend to the words which were highlighted in bold print and glossed in their L1. The involvement index was 1. Participants in group 2 had the same text but the words were deleted from it, each appearing with brief definition, an example sentence and an L1 gloss. The participants were instructed to fill in the blanks with the glosses in the margin. The involvement index was 2. Group 3 only had to write original sentences with the words. The index for this group was assessed to be 3. Based on ILH, it was predicted that group 3 would outperform group 2 which in turn would do better than group 1. The results strongly supported the hypothesis that the involvement load hypothesis can be generalized to low-proficiency learners, though no significant difference was found between the groups on Task 3 and Task 2 about their passive knowledge of the target words.

Kim (2008) also provided empirical evidence for the involvement load hypothesis in a carefully designed study consisting of two experiments. The first experiment addressed the effectiveness of three vocabulary tasks with different levels of involvement index. The second experiment, on the other hand, examined whether tasks with equal involvement load would lead to equivalent initial and later retention of words by 20 adult ESL learners at two different levels of proficiency. In line with other studies, the results

showed that a higher involvement index leads to more effective initial and delayed vocabulary learning. Furthermore, Kim found that identical involvement index in two tasks unfolded similar results for the two L2 proficiencies. Despite the overall support, Folse (2006) reports that his study showed word learning to be more a function of repeated exposure than involvement.

As Keating (2008) and Kim (2008) have recognized, involvement load hypothesis in its current form has not made fine-grained differentiations between the relative load of each involvement factor, nor of the relative importance of the types of tasks (e.g. input versus output), issues that may well have direct impact on word learning. This study limits itself to the second issue, that is, task type.

### **2.3 Research questions**

1. Does involvement load play a part in the initial learning of new vocabulary items?
2. Does involvement load play a part in the retention of new vocabulary items?
3. Is there a difference between input-oriented and output-oriented tasks in initial learning of new vocabulary?
4. Is there a difference between input-oriented and output-oriented tasks in the retention of new vocabulary?

## **3. Method**

### **3.1 Participants**

The participants in this study were 69 female Iranian EFL learners from two all-girls' branches of a private English Language Institute in Mazandaran Province. Each of the branches offered three intact classes for the study. The reason for choosing two branches was that the number of students in each class was around 10, which was assumed to be too small for the study. Hence, two groups were invited to participate in each task and their scores were finally merged into one due to their homogeneity. The participants were largely young adults with an average age of 19. They were



at an intermediate level of proficiency based on the institute's continuous assessment criteria, which we found to be more realistic than a summative test. Their teachers also were of the opinion that the participants were capable of carrying out the required tasks which, thus further corroborates the institute's assessment of the learners' level. Each class was then randomly assigned to one of the three tasks.

### 3.2 Instruments

The instruments in the study were an input task in the form of a text followed by five reading comprehension questions, and an immediate and a delayed posttest. The reading passage was an article selected from *Reading Master* (Liu et al., 2002) which had already been used by Hui-Fang Tu (2003) in a study with a similar purpose. The passage, which consisted of 331 words, was about the suppression of emotions and the potential threats of such behavior to the mental and physical health of human beings. The reason for the selection of this topic was that it was of a general nature and was understandable to the participants, who could relate it somehow to their own personal experiences.

The target words were ten words selected from the above reading passage based on three criteria, i.e., unfamiliarity to the participants, ease of incorporation into a narrative describing one's personal experiences, and ease of supplying a synonym or a definition in the target language as well as an appropriate translation in the participants' native language.

The participants' own teachers, who had wide experience in teaching English to Iranian students in institutes, were consulted regarding the above criteria. They assured the researchers of their appropriateness. The unfamiliarity of the target words to the participants was also ascertained by checking the target words and their respective word families against the list of the words of each of the previously taught as well as their current textbooks. Besides, these ten words were presented to a group of students of the same level who were not supposed to attend the study to ensure the participants' lack of knowledge of these words. The ten target words in the experiments were: *annoy*, *hostile*, *conflict*, *unfortunately*, *suppression*,

*maintain, determination, circumstance, grit, and endure* (four nouns, four verbs, one adjective, and one adverb).

### 3.3 Tasks

Three tasks were used in the study with the target words in bold print to help the participants notice the words (Schmidt, 1994). The first task was a reading comprehension task, which was performed by two of the six groups. The two groups had just to read the text to answer its multiple-choice items. Since the participants had to know the meanings of the target words to answer the comprehension questions, they were told to bring their own dictionary to class and use it when necessary. It has to be noted that all the students in the reading group had already been trained by their teachers before the study began and knew how to use a dictionary. Since it was necessary in this task to use the dictionary to find and figure out the meaning of the polysemous words, all the three involvement components of *need, search* and *evaluation* were present. Hence, based on the involvement load hypothesis (Hulstijn and Laufer, 2001) the involvement index of the first task was 3 (+ (1) need, + (1) search, + (1) evaluation).

The second task was performed by two other groups. The participants had to read the same reading passage with the target words omitted. The target words were placed on top of the page in a random order. Having completed the Gap-fill task, they had to answer the same comprehension questions as the first group. In this task, the need component was moderate, because it was externally induced, i.e., by the task itself. There was no search component since students were provided with the glosses and they did not have to look up the words in a dictionary. In order to fill in the blanks with the correct words, the candidate words provided by the researcher had to be evaluated against one another to determine their contextual appropriacy. The task motivated a moderate amount of evaluation. Based on the involvement load hypothesis the involvement index of the task was 2 (+ (1) need, - (0) search, + (1) evaluation).

In the third task after reading the text and answering the comprehension questions, the participants in the last two groups had to write a composition

of one- to three-paragraphs essay to express their feelings about the incident in the passage using the target words given to them. The students were informed that grammaticality was not an issue, and that they would need to focus on the message they intended to convey. The Persian translations and the English definitions of the target words were provided with the glosses on a separate page.

As to the involvement load, the need and search were the same as those in task two. Since the students were provided with glosses, there was no search component involved, the value of evaluation being, however, higher because the words were to be used in the original context and the participants had to put more effort to create the text. Hence, the involvement load of the task was 3 (+ (1) need, - (0) search, ++ (2) evaluation).

### 3.4 Posttests

To assess initial learning of the target words, an immediate posttest was administered. Also, to determine the retention of the target words, a delayed posttest was administered one week later. The two posttests, which were equal in all respects except the arrangement of the target words, provided the students with the ten target words on a page with the instruction in the participants' first language on top that required of them, following Hulstijn and Laufer (2001), to provide the English equivalents or the Persian translations of these words in writing. It has to be noted that no vocabulary pretest was given to the participants prior to the study to avoid generating any memory traces.

### 3.5 Procedure

To address the above research questions, three different tasks were designed each to be performed in the two branches of the institute (Table 1). Tasks 1 and Task 2 were both input-oriented, with the involvement loads of 3 and 2 respectively. Task three was output-oriented with the involvement of 3. This allowed the researchers to examine the effect of involvement load (experiments 1 and 2) and task type (experiments one and three).

Table 1: Characteristics and sizes of the samples

Institutes	Classes	Number	Tasks
Institute 1	Class 1	12	Task 1: Reading Comprehension + Dictionary(3) *
	Class 2	12	Task 2: Fill in Blanks + Gloss(2)
	Class 3	10	Task 3: Reading Comprehension +Writing a Composition+ Gloss(3)
Institute 2	Class 1	13	Task 1: Reading Comprehension + Dictionary(3)
	Class 2	12	Task 2: Fill in Blanks + Gloss(2)
	Class 3	10	Task 3: Reading Comprehension + Writing a Composition + Gloss(3)

\*: involvement load index

In all the experiments, the worksheets were collected after the completion of the tasks. All the students were given a vocabulary test sheet with a list of the ten target lexical items to provide their meanings (either in Farsi or in English). Due to the nature of the study, incidental learning, the participants were not informed of the upcoming immediate or delayed tests they were supposed to take.

### 3.6 Data Analysis

The two posttests taken by 69 participants were scored by one of the researchers, who assumed 1 point for the correct response and 0 for the wrong one. If an answer was controversial in terms of the degree of semantic approximation, opinions of experienced teachers of the institute were sought for scoring the item. If a learner gave a correct response to an item but had also marked the target word as known to him/her prior to the experiment, the response was scored as zero. So the scores ranged from 0 to 10 for each participant.

There were four research questions in the study each with one dependent variable at two levels, the scores of the participants in the immediate and delayed posttest, and one independent variable, involvement load in the first two questions and task type in the second two. Considering that data met the assumptions suggested by Hatch and Lazaraton (1991) for utilizing a parametric test (interval data, normality of distribution in the sample and the underlying population, and the independence of

observations) separate independent-samples t-tests were conducted to compare the performance of the participants in each two groups. As mentioned above, due to the constraint in terms of the number of students in each class (around 10) the two groups performing the same task were merged to have larger samples, keeping in mind that all the classes were homogeneous in terms of level of proficiency, sex, age, and other possibly related factors.

#### 4. Results

Drawing on Laufer and Hulstijn's (2001) Involvement Load Hypothesis which holds that tasks with a higher involvement load will be more effective than those with a lower involvement load in learning words, the researchers first framed two research questions to compare the findings of the present study with those of Hulstijn & Laufer (2001). The questions and the related responses are presented below one by one.

The first question asked whether involvement load plays a part in the initial learning of new vocabulary items. To address this question, the performances of the participants in the first task, i.e., reading comprehension only, and the second task, i.e., reading comprehension + fill in the blanks, were compared, the reason being that both were input-oriented tasks with different involvement loads. And as detailed in the methods section, the involvement load index for the first and second tasks were 3 and 2 respectively. Hence it was expected that the performance of the two groups be different in favor of the first group. The findings of the study (Table 2) showed that there was a significant difference between the two groups in the immediate posttest ( $t=3.57$ ,  $p<0.05$ ), yet quite contrary to the prediction of the Involvement Load Hypothesis in favor of the second task.

Table 2: The scores of the participants in tasks one and two in the immediate and delayed posttests

Tasks	Number of Participants	Immediate posttest		Delayed Posttest	
		Mean	SD	Mean	SD
Task 1	25	4.16	1.69	2.72	1.33
Task 2	24	6.70	1.49	5.1	1.43

The second question asked whether involvement load play a part in the retention of new vocabulary items. A glance at Table 2 again indicates that as far as the retention of the target words were also concerned, group two, who had completed the reading comprehension + fill in the blanks task, performed significantly better than the first group, who had performed the reading comprehension only task in the delayed test as well ( $t=3.58$ ,  $p<0.05$ ). To sum up, it was found that in both immediate and the delayed posttests, the group which filled in the blanks with the target words performed significantly better than the group which had to look up these words in the dictionary to answer the multiple choice comprehension questions.

Unlike the first two research questions, which were a partial replication of Hulstijn & Laufer's (2001) study on the role of involvement load, the second two research questions were genuine questions, which addressed the role of task type, investigating the effect of task type on learning vocabulary. The third question asked whether there is a difference between input-oriented and output-oriented tasks in the initial learning of new vocabulary. As mentioned in the method section, tasks one and three were of equal task-induced involvement load, that is, 3 in both cases, the former being input-oriented and the latter output-oriented.

Table 3: The scores of the participants in input-oriented and output-oriented tasks in immediate and delayed posttests

Tasks	Number of Participants	Immediate posttest		Delayed Posttest	
		Mean	SD	Mean	SD
Task 1	25	4.16	1.69	2.72	1.33
Task 3	20	7.6	0.94	6.15	1.18

The above table displays significant differences between the mean scores of the participants in the two tasks of 1 (i.e., reading comprehension only) and 3 (reading comprehension + writing composition) in favor of task 3 in the immediate posttests ( $t=3.57$ ,  $p<0.05$ ). In other words, those participants who had performed the output-oriented task obtained better results than those who had completed the input-oriented task as far as initial

learning of vocabulary is concerned.

Finally, the fourth question asked whether there is a difference between input-oriented and output-oriented tasks in the retention of new vocabulary. As to the retention of the target vocabulary also, the analyses of the data through t-tests revealed that there was a significant difference between the group that had completed the input-oriented task and the one that had learned vocabulary through the output-oriented task, of course, in favor of the latter ( $t=3.58$ ,  $p<0.05$ ). This indicates that, everything else being equal, output-oriented tasks are far more conducive to learning vocabulary than input-oriented tasks.

## 5. Discussion

The first two research questions put the predictions of the involvement load hypothesis to empirical test. Unlike previous studies (Keating, 2008; Kim, 2008; Huljistin and Laufer, 2001) which rather unambiguously endorsed the hypothesis, our results paint a very different picture. Based on the hypothesis, it was anticipated that initial and delayed word retention to be a function of involvement index, i.e., Task 2 to be less effective than Task 1. However, based on the data analysis, it turned out that it was the other way round. One possible explanation for this is that the numerical values given to the motivational and cognitive elements, which in turn yield the involvement index, may not carry the same weight, or may have been roughly quantified. This is very much in line with Kim's (2008) argument. According to Kim "it is possible that all three components might not be equal in contributing to vocabulary learning (p. 313)." Along the same line, it might be claimed from the results of this study that each individual component (in this case the *evaluation* component) may have a higher variable involvement load. In Laufer and Huljistin (2001) the involvement index for evaluation for both "a comparison of a given word with other words, a specific meaning of a word with other meanings (p.14)" is considered equivalent, i.e., 1. The results of this study suggest that not all input-oriented tasks demand equal cognitive processing. In Task 1 the task-induced evaluation incidentally required the participants to locate the

particular meaning of the polysemous lexical item from a very finite set and to answer the multiple questions, an activity whose depth of processing seems to be much less than Task 2, in which learners had to weigh the glosses against one another to single out the ones that most appropriately fitted the blanks, a cognitive process which may well account for their superiority. On the whole, the study suggests that the grading of components of involvement load hypothesis should be further fine grained to be proportionate to the depth of processing (in case of this study '*evaluation*') that a task demands.

The third and the fourth questions sought to address the probable link that might exist between the involvement hypothesis and task type. Based on the hypothesis we expected no significant difference between word retention in tasks 1 and 3, as both were equivalent in involvement index. After all, according to the Huljistin and Laufer (2001: 552) “the Involvement Load hypothesis does not predict that any output task will lead to better results than any input task. It predicts that higher involvement induced by the task will result in better retention, regardless of whether it is an input or an output task.” Furthermore, studies (Kim, 2008; Keating, 2008) had already shown that tasks of equal load will lead to similar results irrespective of proficiency level. The results of this study, however, suggest that despite equivalency of involvement index in Task 1 and 3, word retention was statistically different, a finding that is at odds with the prediction of the involvement load hypothesis. Two different explanations may be adduced for this. The first explanation has to do with the task type. As Task 3 had students write connected discourse, the act of production itself, which demands deeper cognitive effort (Swain, 1985, 1995, 2000) might have contributed more to word retention than the mere reading of the text. In other words, what contributes to word retention is not merely a product of deliberate manipulation of variable elements (*need, search and evaluation*) in task, irrespective of its type, rather other elements such as task type may be equally important. Another explanation for the result might be related to the materials and procedures used for this study. Unlike previous studies, Task 3 in this study was more elaborate (and possibly



more natural) in that the composition part followed after reading the passage and answering the related comprehension questions. The previous studies only included a list of words with which the participants were to write original sentences. Hence it seems to be a plausible explanation to suggest that in this study word retention was a product of repeated exposure. However, as the participants did not have repeated exposure to the target words in original texts, there is ample room for skepticism.

## 6. Conclusion

In this paper, it was suggested that although the involvement load hypothesis may provide a reasonable point of departure for vocabulary instruction, it has long way to go before it achieves its full potential. To start with, the constituent components yielding the involvement index may well need to be more rigorously defined. This study has shown cases where the involvement load elements had been configured in line with their theoretical operationalization, that did not live up to the predictions of the hypothesis. Secondly, the involvement index may well not function independently of the task type for vocabulary instruction. That is, the processing load brought to bear by task type may well affect word retention, a point needing further empirical studies.

Despite the challenges facing the hypothesis, it provides L2 teachers with a good foothold in vocabulary instruction as it supplies them with a theoretical, other than impressionistic, tool in vocabulary instruction. However, the findings of this study suggest that caution must be exercised by practitioners as to the synergistic effect of the three element of the construct of involvement. Hence, the findings of this study suggest that further complementary studies be conducted to improve upon the hypothesis, especially on the effect of the evaluation element on task type, as the two seem to be more effective in vocabulary learning.

## References

- Atkinson, R. and Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In Spence, K. and Spence, J. (Eds.), *The*

- psychology of learning and motivation: advances in research and theory*, New York: Academic Press, 2, 89-195.
- Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 85, 139-152.
- Craik, F. I. M. and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Eysenck, M. W. (1978). Levels of processing: A critique. *British Journal of Psychology*, 69, 157-169.
- Eysenck, M. W. (2004). *Psychology: An international perspective*. Hove: Psychology.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Journal*, 40, 273-293.
- Hatch, E. M. and Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York, NY: Newbury House Publishers.
- Hui-Fang, Tu. (2003). Effects of task-induced involvement on incidental vocabulary learning in a second language. Available: <http://www.hss.nthu.edu.tw/~fl/thesis/tesol/905258.pdf>.
- Hulstijn, J. H. and Laufer, B. (2001). Some empirical evidence for involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539-558.
- 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.
- Laufer, B. and Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, 22 (1), 1-26.
- Liu, W., Yun, J., Chen, N., Tsao, S., McCormack, S., and Wheeler, N.

- (2002). *Reading master*. Taipei: AMC Publishers.
- Nagy, W., Herman, R. and Anderson, R. (1985). Learning words from context. *Reading research quarterly*, 20, 233-253.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behaviour*, 16, 151-171.
- Nunan, D. (1999). *Second language teaching and learning*. Boston: Heinle & Heinle Publishers.
- Olson, M. H. and Hergenhahn, B. R. (2009). *An introduction to theories of learning*. Prentice-Hall.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11. 2, 17-45.
- Schmitt, N. (2008). Review article: instructed second language vocabulary learning. *Language Teaching Research*, 12 (3), 329-363.
- Solso, R. L. (1988). *Cognitive psychology* (2nd 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. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition*. Rowley, MA: Newbury House, 235-253.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook and B. Seidlhofer (Eds.), *Principles and practice in applied linguistics*. Oxford: Oxford University Press.
- VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition* 12. 3, 287-301.
- VanPatten, B., Williams, J., Rott, S. and Overstreet, M. (2004). *Form-meaning connections in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

