

Zargham Ghabanchi

Ferdowsi University of Mashhad, Iran

Mohammad Davoudi

Sabzevar Tarbiyat Moalem University, Iran

Z. Eskandari

Sabzevar Tarbiyat Moalem University, Iran

Vocabulary learning through input and output tasks: Investigating the Involvement Load Hypothesis

Abstract. Based on Hulstijn and Laufer (2001), a study was carried out consisting of two experiments investigating the involvement load hypothesis in vocabulary learning. Experiment 1 compared the performance of 88 Iranian adult English as foreign language (EFL) learners at the same proficiency level (i.e., high intermediate) to ascertain the effectiveness of three vocabulary tasks with different levels of task-induced involvement. Experiment 2 investigated whether three different tasks which are of different types (input vs. output) but hypothesized to represent the same level of task-induced involvement would result in equivalent *initial* learning and *retention* of target words by 56 Iranian adult EFL learners at the same level of proficiency. The results of Experiment 1 showed that the higher level of learner involvement during the task promoted more effective *initial* learning and better *retention* of the new words. The findings of Experiment 2 indicated that when different tasks had the same involvement load, they resulted in similar amounts of *initial* vocabulary learning and *retention* of new words. The results of the two experiments are discussed in light of the involvement load hypothesis.

Keywords: involvement load hypothesis, processing depth theory, task-induced involvement, foreign language vocabulary acquisition, English as a foreign language

Introduction

Vocabulary learning is considered central to language acquisition, whether the language is first, second or foreign language (Decarrico, 2001; Krashen, 1989; Mobarge, 1997; Poulisse & Schils, 1995). There is a general agreement among vocabulary specialists that lexical competence is the

very heart of communicative competence, the ability to communicate successfully and appropriately (Coady & Huckin, 1997).

Different approaches to teaching and learning vocabulary have been proposed in literature of which “incidental” and “intentional” learning could be named. Hulstijn and Laufer (2001) define intentional vocabulary learning as “the activity aimed at committing lexical information to memory” (p.11). On the other hand, an incidental learning situation is a situation in which individuals process new information without the intention to commit this information to memory. Worded differently, incidental vocabulary learning refers to “the learning of vocabulary as a by-product of any activity not explicitly geared to lexical learning” (p.10).

A typical and well-known proponent of incidental vocabulary learning is Krashen (1989), who, in the context of his *input hypothesis*, argues that we acquire vocabulary and spelling through exposure to *comprehensible input*, that is; “by understanding language that is ‘a little beyond’ our current level of competence”(Krashen, 1981, p.103). In his words, comprehension is at the heart of language acquisition and production is only a sign of second language acquisition that has already taken place (Krashen, 1981, 1989).

Swain (1985), by acknowledging the invaluable role *comprehensible input* serves in second language acquisition (SLA), argues that it is not sufficient for learners of English as a foreign (EFL) or second language (ESL) to fully develop their second/foreign language (L2) proficiency. According to Swain what these learners need is not only comprehensible input but *comprehensible output*. The construct holds that when learners run into communication difficulties, they try to make their output more coherent, appropriate and precise. The assumption is that this process would contribute to language learning by making learners notice the gap between their interlanguage and language as it is used by native speakers and discover what they

can do and cannot do.

Another crucial concept underlying most of the research on incidental vocabulary learning is the *depth of processing framework* proposed by Craik and Lockhart (1972). It holds that “the memory trace can be understood as a byproduct of perceptual analysis and that trace persistence is a positive function of the depth to which the stimulus has been analyzed”(p.671). Based on this framework the deeper the processing of a stimulus is, the traces in memory will be more elaborate, longer lasting, and stronger. The hypothesis suggests that the retention of information is determined by the depth to which it is processed rather than the length of time it is held in the primary memory. They have also posited several levels of processing depth. For example processing the semantic features of a lexical item (e.g., meaning) is supposed to occur at a deeper level than its structural features (e.g., orthography). In other words, tasks which require learners to process the meaning of words lead to better word retention.

Purpose of the Study

The present study was designed to test particular predictions of the involvement load hypothesis in English as a foreign language (EFL) setting. The study was composed of two experiments. The first, Experiment 1, was a partial replication study of Hulstijn and Laufer’s (2001) Dutch/Hebrew experiments. The only difference was that in their study the students were of different proficiency levels, but in the present study the participants were of the same level of language proficiency (high intermediate). The purpose of Experiment 1 was to examine how different levels of task-induced involvement affect the initial learning and retention¹ of target words by L2 learners of the same level of proficiency. Experiment 2 was also carried out in an EFL setting with learners of the same level of proficiency (high intermediate). The purpose of Experiment 2 was to investigate Laufer and Hulstijn’s (2001) claim that no particular task type –

be it input or output – is considered superior or more effective and that the only determining factor in task efficacy is the degree of involvement that a task induces. In both experiments, time-on-task, as claimed by Laufer and Hulstijn (2001), was considered as an inherent feature of the task.

Experiment 1

Research Questions

1. Does the level of task-induced involvement affect the *initial vocabulary* learning of EFL learners of the same proficiency level when three tasks with different levels of involvement are administered?
2. Does the level of task-induced involvement affect the *retention* of new vocabulary of EFL learners of the same proficiency level when three tasks with different levels of involvement are administered?

Based on the above questions the following hypotheses are postulated:

H0a: The level of task-induced involvement does not affect the *initial vocabulary* learning of EFL learners of the same proficiency level when three tasks with different levels of involvement are administered.

H0b: The level of task-induced involvement does not affect the *retention* of new words of EFL learners of the same proficiency level when three tasks with different levels of involvement are administered.

Method

Participants

A total of 120 EFL adult learners, whose native language was Persian, participated in the study. The sample included both males and females ranging from 19 to 25 years old. Participants were

students of English literature at Sabzevar Teacher Training University in Iran. They were grouped into three levels of proficiency based on their paper-based Test of English as a Foreign Language (TOEFL) score and those placed in the high intermediate level were selected for the study. All of the participants were from five intact classes and the data were collected during their regularly scheduled class periods. During each class meeting participants of high intermediate level of proficiency were randomly assigned to one of the three experimental tasks. Others were given a Reading comprehension task as placebo.

Materials

Target words. A total of ten target words were selected according to the following criteria: First, 30 students who were representative of the sample selected for the study, were asked to choose unknown words from a list of 20 words selected from the reading material. Based on their responses, 10 target words that were unfamiliar to all students were selected for the study. The target words consisted of six nouns, three verbs and one adjective: *risible, betrothal, coven, guffaw, sortie, smattering, quandary, berate, whiff, and slop.*

The reading text. The reading passage was adopted from an article in a reading comprehension book *The World of Words* (Richeck, 1993). The passage was about the origins of superstitions. The text contained 551 words and was selected on several grounds. First, the participants were supposed to have some general idea of the topic and second, they were capable of writing about their personal experience pertinent to the topic. Ten multiple-choice reading comprehension questions were also prepared based on the text, the completion of which required understanding of the target words. The reading passage was modified, as were the comprehension questions. The aim of controlling the vocabulary was to bring the majority of the words within the learners' experience without oversimplifying the text. In addition, reducing the number of unknown words

in the text frees up the amount of cognitive space required to attend to the message (Joe, 1998). Thus, as additional resources are made available during text processing, the forging of stronger form-meaning connections is made possible, such that target words may be retrievable at a later time (Craik & Lockhart, 1972; Craik & Tulving, 1975; Laufer & Hulstijn, 2001; Nation, 2001). Another criterion for modifying the text was the number of occurrences for each target word. The passage was revised in such a way that all target words would appear only once.

Tasks. Three tasks with different involvement loads were set to serve the purpose of study in Experiment 1.

Glossing Task: Learners performing this task were provided with a text and ten multiple-choice comprehension questions based on the reading passage. These questions either incorporated some target words or paraphrased the original sentences in which these target words occurred. The words were glossed by providing their L1 translation and their L2 explanation on the margin of the text. The task of the students was to read the text and answer the ten comprehension questions. In terms of involvement load, the task induced moderate need, but neither search nor evaluation. Its involvement index was 1.

Fill-in Task: Ten target words were deleted from the text, leaving ten gaps numbered 1-10. The ten target words, along with five words that had not appeared in the original text, were printed in random order as a list on a separate page, with their L1 translation and their L2 explanation.

Composition Writing Task and Definition: were also used as testing procedure

Procedure

The experiment was conducted over a two-week period. The treatment and immediate vocabulary test were administered on the same day and the delayed posttest two weeks later. As Hulstijn and Laufer (2001), time-on-task was regarded as an inherent property of a task. Since

the study aimed to investigate the retention effects of tasks, no attempt was made to control for time-on-task. In order to estimate the time spent on each task, the participants were asked to show the starting time and finishing time on their worksheets. Glossing, Fill-in, and Composition writing tasks took approximately 12, 17, and 25 minutes respectively. Upon the completion of the tasks, the worksheets were collected and the students were unexpectedly given an immediate posttest designed to measure their initial vocabulary learning. They were also asked to indicate whether they had known the words prior to the task in order to check for the preknowledge of the target words. If a participant checked more than two words as previously known, the data collected from that student were eliminated from the analysis. Two weeks later, participants received the delayed posttest which had the same format but which displayed the test items in a different order to measure their retention of vocabulary knowledge. The tests were then scored by two expert raters.

Analysis

Each research question examined one dependent variable: scores on the immediate and delayed posttests. Both research questions had the same independent variables: level of task-induced involvement. In order to test the two research questions, the data were analyzed using SPSS version 16. Two one-way ANOVAS were conducted to analyze the results in order to address each of the two research questions. The alpha level was set at .05 when significant results were found.

Results

Descriptive statistics for the immediate and delayed posttests are displayed in Table 1. Overall the descriptive statistics indicated that the Composition group acquired the highest mean scores on the VKS in both the immediate and delayed posttests.

Table 1 Results of descriptive statistics for immediate and delayed posttests in Experiment 1

tasks	N	Immediate		Delayed	
		M	SD	M	SD
Glossing (1)	32	27.25	6.46	21.62	4.75
Gap-fill (2)	25	28.00	6.51	22.20	3.79
Composition writing (3)	31	34.58	8.50	25.65	7.56

The scores of each posttest were then submitted to a one-way ANOVA to determine if the differences were significant (Table 2 and Table 3).

Table 2 ANOVA on retention scores of the immediate posttest in Experiment 1

College	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	992.270	2	496.13		9.39.000
Within Group	4487.548	85	52.79		
Total	5479.818	87			

* $p < .05$

The ANOVA results for the first research question in Experiment 1 indicated a significant difference among the means [$F(2, 85) = 9.39; p < .001$]. The first two tasks resulted in little retention; the Composition writing group fared significantly better.

The Scheffe posthoc test revealed that the mean score of the composition writing group was significantly different from the mean scores of the Glossing group ($p = .001$) and the Gap-fill group ($p = .005$), but that the means of the latter two groups did not significantly differ from each other ($p = .92$). So, based on the results obtained, the first null hypothesis, that task type does not have any effect on the amount of *initial* vocabulary learning, is rejected.

Table 3 ANOVA on retention scores of the delayed posttest in Experiment 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	289.72	2	144.86	4.45	.014
Within Group	2764.59	85	32.52		
Total	3054.31	87			

* p<.05

As with the first research question, the results for the second research question in Experiment 1 revealed a significant difference among the means [$F(2, 85) = 4.45$; $p < .05$]. The Scheffe posthoc revealed that the mean score of the Composition group was significantly different than the mean score of the Glossing group on the delayed vocabulary posttest ($p = .02$), but the mean of Composition writing group did not differ significantly from that of the Gap-fill group ($p = .08$). The means of the Glossing and Gap-fill group were not significantly different either ($p = .93$) as it was the case with the immediate posttest scores. Thus, based on the observed results, the second null hypothesis, that the level of task-induced involvement does not affect the *retention* of new words, is also rejected.

Table 4 Results of Paired *t* test for the tasks in Experiment 1

Pairs	M	SD	<i>t</i>	df	Sig.
Glossing T ₁ & T ₂	5.625	4.492	7.084	31	.000
Gap- fill T ₁ & T ₂	5.800	4.655	6.230	24	.000
Composition T ₁ & T ₂	8.935	5.428	9.166	30	.000

Note. T₁ and T₂ stand for the immediate and delayed posttests respectively.

Comparing the means of the immediate and delayed posttests for each task (see Table 1), a decline can be observed for each task in the delayed posttest. Hence, the scores of the immediate and delayed posttests of each task were submitted to a Paired *t* test to see if the mean differences were significant. As Table 4 reveals, the mean differences of all tasks are significant at $p < .001$.

Experiment 2

Experiment 2 examined how three tasks with the same theoretical level of task-induced involvement affected the *initial* learning and *retention* of target words by L2 students with the same level of proficiency. According to the involvement load hypothesis, three tasks that induce the same amount of involvement should promote similar results in vocabulary acquisition. In Experiment 2 the efficacy of three tasks (Dictionary use, Sentence writing, and Composition writing) with the same involvement index of three were compared.

Research questions

Experiment 2 was guided by the following research questions.

1. Does the type of task have an effect on the amount of *initial* vocabulary learning by EFL students of the same level of proficiency when three tasks with the same level of involvement are administered?
2. Does the type of task have an effect on the retention of new vocabulary by EFL students of the same level of proficiency when three tasks with the same level of involvement are administered?

Based on the above questions the following hypotheses are postulated:

H0_a: Task type does not have any effect on the amount of *initial* vocabulary learning by EFL students of the same level of proficiency when three tasks with the same level of involvement are administered.

H0_b: Task type does not have any effect on the *retention* of new vocabulary by EFL students of the same level of proficiency when three tasks with the same level of involvement are administered.

Method

Participants

The target population in Experiment 2 was similar to those participated in Experiment 1.

Materials

Target vocabulary. The same 10 target words from Experiment 1 were also used in Experiment 2:

The reading text. The same reading text used in Experiment 1 for Glossing and Fill-in tasks was also used here for the Dictionary use task.

Testing material and scoring procedure. The vocabulary knowledge measures for Experiment 2 were the same as for experiment 1. Paribakht and Wesch's (1997) VKS was used to measure the participants' initial learning and retention of new vocabulary knowledge.

Tasks. three different tasks with the same involvement indexes were used in the experiment of the tasks, the Dictionary use task was considered as an input task and the two other ones (i.e., Composition writing and Sentence writing) were deemed as the output tasks since they involved production.

Composition writing task:

Dictionary use task: Learners performing the task were provided with a text and ten multiple-choice comprehension questions based on the reading passage.

Sentence Writing Task: Participants assigned to this task were provided with a list of the ten target words.

Procedure

The procedures applied to conduct the treatments and administer the immediate and delayed posttests were identical to those in Experiment 1.

Analysis

The same statistical procedures as those used in Experiment 1, were used to analyze the data in Experiment 2.

Results

Descriptive statistics for the immediate and delayed posttests are displayed in Table 5. As Table 5 shows, there is little difference among the three tasks in both immediate and delayed posttests. The scores of each posttest were then submitted to a one-way ANOVA to determine if the differences were significant (Table 5 and Table 6).

Table 5 Results of descriptive statistics for immediate and delayed posttests in Experiment 2

Tasks	N	Immediate		Delayed	
		M	SD	M	SD
Glossing (1)	23	29.04	8.76	22.78	6.86
Gap-fill (2)	33	32.91	7.91	23.06	5.78
Composition writing (3)	31	34.58	8.50	25.65	7.56

Note. Each task is followed by the involvement index in parentheses.

Table 6 ANOVA on retention scores of the immediate posttest in Experiment 2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	414.49	2	207.24	2.96	.057
Within Group	5867.23	84	69.84		
Total	6281.72	86			

The first research question examined whether there was an effect of task type on the amount of *initial* vocabulary learning by EFL students at the same level of proficiency when three tasks with the same involvement were administered. As Table 6 shows, no significant results were

revealed ($\rho = .057$). This indicated that there was no significant difference among the tasks in the immediate posttest. Hence, the first null hypothesis in Experiment 2 is accepted.

The second research question asked whether there was an effect of task type on the amount of retention of new vocabulary by EFL students at the same proficiency levels when three tasks with the same level of involvement were completed. As with Research Question 1 in Experiment 2, the results of the one way ANOVA revealed no significant results. In other words, no significant differences were observed between the three tasks.

Table 7 Results of Paired *t* test for the tasks in Experiment 2

Pairs	M	SD	<i>t</i>	<i>df</i>	Sig.
Dictionary T ₁ & T ₂	6.26	5.16	5.81	22	.000
Sentence T ₁ & T ₂	9.84	6.01	9.41	32	.000
Composition T ₁ & T ₂	8.93	5.42	9.16	30	.000

Note. T₁ and T₂ stand for the immediate and delayed posttests respectively.

Since like Experiment 1, the retention mean score of each task in the delayed posttest was less than that in the immediate posttest), the retention scores of each task in the immediate and delayed posttests were submitted to a Paired *t* test (Table 7). Table 8 reveals that the mean difference between the immediate and delayed posttests for each task is significant at $\rho < .001$.

Discussion

The purpose of Experiment 1 was to examine how different levels of task-induced involvement affected the *initial* learning and *retention* of target words by EFL learners who were of the same level of proficiency. Experiment 1 was partially a replication of Hulstijn and Laufer's (2001) original study. The results of the first research question in Experiment 1 are in harmony with those of Dutch-English experiment in Hulstijn and Laufer (2001) and those of Kim (2008) in that

Composition writing task yielded higher retention than Glossing and Gap-fill tasks but that Gap-fill task did not produce retention significantly higher than Glossing task.

By the same token, learners in the current study acquired words more effectively through a task that required a higher level of involvement load. The Composition group yielded higher scores on the immediate posttest than the Glossing and Gap-fill group. However, the Gap-fill group did not perform significantly better than the Glossing group.

One possible explanation for the observed results might be that the extent to which different degrees (moderate and strong) of each individual component (need, search, and evaluation) that might contribute to an overall involvement load, might not be the same, at least not for the initial vocabulary acquisition (Kim, 2008). Hulstijn and Laufer (2001) deemed every component equally effective when specifying the involvement index for vocabulary tasks; however, as Kim clearly states, it might be the case that strong evaluation induces much greater involvement in processing a word than the moderate evaluation and the other two components.

In answering the second research question, which investigated the *retention* of new vocabulary knowledge, Experiment 1 again provided partial support for the Involvement Load Hypothesis. In the delayed posttest, participants who performed the Composition writing task scored significantly better than those who performed the Glossing task, but there were no significant differences between the Composition writing and Gap-fill tasks and between the Glossing and Gap-fill tasks. Keating (2008) also obtained results similar to the current study. Furthermore, in accordance with the results of most previous studies on incidental vocabulary acquisition (e.g., Hulstijn & Laufer, 2001; Keating, 2008; Watanabe, 1997), the present study reports a significant decline in the performance of all the three groups in Experiment 1, from the immediate posttest to the delayed posttest; that is, after a period of two weeks.

Experiment 1 partially supported the claim that words which had been initially processed with higher involvement load were retained better than the words that had been processed with lower involvement load, a finding constant with the claim that elaborate processing fosters retention (e.g. Baddeley, 1990; Craik & Lockhart, 1972).

Experiment 2 examined whether three tasks (i.e. Composition writing, Sentence writing and Dictionary use) claiming to involve the same level of task-induced involvement but are of different types (input and output) would have similar effects on the *initial* learning and *retention* of target words by EFL learners of the same level of proficiency. The results suggested that the three tasks with identical involvement loads were equally effective in promoting both the *initial* learning and *retention* of new words and thus lend strong support to Laufer and Hulstijn's (2001) claim that the effectiveness of a task is determined by the involvement load it induces irrespective of whether the task is input or output oriented. As Craik and Tulving (1975) notes, a determinant factor of retention is "the nature of encoding operation" (p. 278). This equivalent involvement load for the three tasks led to equal levels of performance for the three groups in *initial* vocabulary learning and in *retention* of the target lexical items.

Overall, Experiment 1 partially supported the hypothesis the words which are processed with higher involvement load will be retained better than words that are processed with lower involvement load. The results of Experiment 2, however, fully supported the claim that any output task will lead to better results than any input task and that the only determinant factor is the involvement induced by the task.

Conclusions

This study involves both theory building and theory testing. It is theory building in the sense that a theory of vocabulary acquisition is far from being comprehensive, as stated by many scholars (Meara, 1998), and that this study can provide some contribution to the building of this theory. It

is theory testing in the sense that it is basing its hypotheses on the predictions of a certain theoretical framework (the Involvement Load Hypothesis) and finding out how successful this framework can be in predicting the outcomes.

The results of the two experiments presented in this study generally corroborated Hulstijn and Laufer's (2001) findings. Experiment 1 aimed at investigating whether higher task-induced involvement leads to more effective *initial* learning and *retention* of new vocabulary knowledge. It partially supported the Involvement Load Hypothesis suggesting that higher involvement induced by the task resulted in more effective *initial* vocabulary learning and better *retention* of new words. Experiment 2 was designed to scrutinize the effect of three different tasks with the involvement indexes on *initial* learning and *retention* of the target words. The results revealed that tasks were equally beneficial for *initial* learning and *retention* of target words when their involvement loads were the same; thus providing strong support for the hypothesis. However, for the comparison of the findings of Hulstijn and Laufer's original study and the current study, the current study should not be considered a strict replication because the two studies did not use the same test measurements. As a result it is essential to recognize the characteristics of different vocabulary tests used in each study when applying the results to the Involvement Load Hypothesis.

References

- Baddeley, A. (1990). *Human memory: Theory and practice*. Hove, England: Lawrence Erlbaum.
- Baddeley, A. (1999). *Essentials of human memory*. London: Psychology Press.
- Coady, J., & Huckin, T. (eds.) (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. New York: Cambridge University Press.
- Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671- 684.
- Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in Episodic memory. *Journal of Experimental Psychology: General*, 104 (3), 268-294.
- Decarrico, J. (2001). Vocabulary learning and teaching. In Celce-Murcia (Ed.), *Teaching English as a second or foreign language*. Boston: Heinle and Heinle.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.
- Hulstijn, J. (2003). Incidental and intentional learning. In C.J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition* (pp. 349-381). Blackwell Publishing.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539 - 558.
- Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19, 357-377.
- Keating, G. (2008). Task effectiveness and word learning in a second language: the involvement load hypothesis on trial. *Language Teaching Research*, 12(3), 365-386.
- Krashen, S. (1981). *Second language acquisition and second language learning*. California: Pergamon Press Inc.
- Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73, 440- 464.
- Meara, P. (1998). Second language vocabulary acquisition: a rationale for pedagogy. *Applied Linguistics*, 19(2), 289-292.
- Mobarge, M. (1997). Acquiring, teaching and testing vocabulary. *International Journal of Applied Linguistics*, 7(2), 201-222.

- Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
- Poulisse, N., & Schils, E., (1995). The influence of task and proficiency related factors on the use of compensatory strategies: a quantitative analysis. In B. Hardely (Ed.), *Lexical issues in language learning*. Amsterdam: John Benjamin Publishing Company.
- Swain, M., (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden, (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19, 287-307.

Contact:

ghabanchi@um.ac.ir