The Role of Language Glossing in a Rooted Theory: The Involvement Load Hypothesis

Hassan Soleimani

Department Of Applied Linguistics, Payame Noor University, Iran

PO box 19395-3697 Tehran

E-mail: arshia.soleimani@gmail.com

Mahboubeh Rahmanian (corresponding author)

Payame Noor University, Iran

PO box 19395-3697 Tehran

E-mail: rahmanian74@yahoo.com

Abstract

This study builds on the innovative construct of *task-induced involvement*. It investigates different first and second language glossing in "involvement load hypothesis". In order to do our study, 66 learners from two different English institutes were classified into two major high and low proficient groups based on Nelson Proficiency Test. Among them 22 low proficient students were randomly assigned to two different subgroups in order to complete two different types of tasks with different language glossing. The behaviors of the two groups in the immediate and delayed posttests did not confirm the predictions of involvement load hypothesis. Although the second task had a higher involvement index, no sign of superiority was observed in its performance. The study suggests that the predictions of involvement load hypothesis need to be more finely grained especially when the tasks have different language glossing.

Keywords: involvement load hypothesis, task-induced involvement, involvement index, L2 glossing, L1 glossing, retention, evaluation

1. Introduction

Finding out an effective vocabulary learning strategy has long been the focal point of many researches. Although there are many ways of learning it, we can not find unanimous beliefs in this area. We can consider incidental vocabulary learning as one these challenging strategies since there is not enough agreement in its usefulness. Understandably, Krashen (1989) knows extensive reading (i.e., the origin of incidental vocabulary learning) as the reason of repertoire of vocabularies, but Laufer (2001) does not consider it so much important. Besides, Laufer and Hulstijn (2001) believe that "in the majority of incidental vocabulary acquisition studies, learners are typically required to perform a task involving the processing of some information without being told in advance that they will be tested afterwards on the recall of all the words in the list" (p.10). In other words, they approached incidental learning in a new manner in which participants are tested surprisingly, without prior informing them. They suggested their "involvement load hypothesis" as one of the branches of incidental learning which states that by changing the amount of its constituent parts, cognitive (*need*, *search*) and motivational (*evaluation*) factors, the retention of unfamiliar vocabulary items will be affected. This theory like the previous ones has many dark points that has not yet been investigated comprehensively like first or second language (L1 or L2) glossing (Keating, 2008), and regarding the same value to its components (i.e., need, search, evaluation; Kim, 2008). The present study is determined to reveal these points in involvement load hypothesis.

2. Literature Review

Won (2008) defines vocabulary knowledge as "the every dimension of complex word knowledge related to comprehension" (p.11). This knowledge is of so much importance that other things like selecting the best techniques of vocabulary instruction are relied on it. There is a strong correlation between vocabulary knowledge and successful L2 learning (Huckin & Bloch, 1993; as cited in Won, 2008). Up to date, many researchers have tried to investigate the role of vocabulary in other fields of language. Schmitt, Schmitt, and Clapham (2001) remind us that vocabulary is "the building block of language" (p. 53). Although many researchers (e.g., Knight, 1994; Groot, 2006) accept it as the principal element of language proficiency, we are yet in the darkness of ambiguities of learning it.

Incidental learning as one of the strategies of learning vocabularies was established by Krashen (1989) who draw our attention to extensive reading and claims that a major part of our vocabulary store is in debt of it. Committing the words to memory without conscious intention and picking up unfamiliar words are the key factors in another definition by Hulstijn (2011) in incidental vocabulary learning. On the other hand, Laufer and Hulstijn (2001) have seen incidental learning in the light of testing the participants surprisingly. They call it as involvement load hypothesis which consists

of need, search and evaluation factors. They argue that by enhancing the amount of its cognitive and motivational components, the retention of unfamiliar target vocabulary items will become better.

Evaluation is one of the cognitive factors of the involvement index. The main issue in this component is context which can specify the appropriate target word in terms of its form, usage, collocations, and meaning or use. Understandably, this component 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" (Hulstijn & Laufer, 2001, p.14). Evaluation can be "moderate" or "strong": If a task requires learners to only assess the appropriate word for a specified context it has "moderate evaluation", but if in addition it induces them to make a sentence with it "strong evaluation" occurs. Furthermore, Crookes and Schmidt (1991, as cited in Laufer & Hulstijn, 2001) know the need component as part of the motivational construct which Cheng (2011) considers it to be the most suspicious one. Walsh (2009) knows it as the driving force behind task completion which in accordance with (Laufer & Hulstijn, 2001) can be either extrinsic like the need to use a word in a sentence because of teacher's request, "moderate need", or intrinsic like when it is self-imposed by the learner himself or herself, "strong need". Search, which is another cognitive factor in involvement index, is concerned with noticing to the form-meaning relationship of the target words (Schmidt, 1994, 2000). Laufer and Hulstijn (2001) refer to this factor to state the importance of finding the form or meaning of unknown L2 vocabularies via a main source of information such as a dictionary or teacher.

There can be found some researches in the literature that tried to better investigate the Laufer and Hulstijn's (2001) innovative construct. Even though, some researchers (i.e., Craik & Lockhart, 1972; Craik & Tulving, 1975; Laufer & Hulstijn, 2001; Nation, 2001) hold that glossing induces a better relationship between form and meaning and in effect a better retention, the role of its language has not yet been approved thoroughly in the hypothesis. Keating (2008), for instance, investigated the effect of L1 and L2 glosses in different types of tasks in involvement load hypothesis to see their effects on vocabulary acquisition.

2. 1 Research Questions and Hypotheses

The study is intended to address the following questions:

Research Question 1:

Does using L2 glossing in involvement load hypothesis have any effect on enhancing vocabulary acquisition of Iranian EFL students?

RQ2: Does using L1 glossing in involvement load hypothesis have any effect on enhancing vocabulary acquisition of Iranian EFL students?

In order to gain access to more or less convincing findings to remove the pertinent ambiguities, the following null hypotheses were formulated.

Null Hypothesis 1:

Using L2 glossing in involvement load hypothesis has no effect on enhancing vocabulary acquisition of Iranian EFL students.

Null Hypothesis 2:

Using L1 glossing in involvement load hypothesis has no effect on enhancing vocabulary acquisition of Iranian EFL students.

3. Method

3. 1 Participants

The population of the present study was selected from 66 Iranian English as foreign language (EFL) male and female students between 19- 25 years old. These intermediate students were from two branches of English Language Institutions in Isfahan, Iran. Each institution provided three intact classes that had 11 students who were in the same classes for three English terms and in effect studied the same core and supplementary materials. This investigation was conducted during normal class time in about a six-week period of June-August 2013 term. In order to have a more homogeneous sample population Nelson Placement Test was administered to the 66 students. Based on their scores, they were classified in to two high and low proficient groups. Afterwards, the first 22 students of the low proficient group were selected as the sample population of our study and in turn were assigned to two different groups to do two different tasks.

3.2 Instruments

Some materials were used for collecting data to conduct the research investigation: Nelson Placement Test and a reading passage with five reading comprehension questions.

3. 2.1 Nelson Placement Test

Since two homogeneous sample groups were needed for conducting the study, the 66 students of the two institutions were given the second version (intermediate) 200 A of Nelson Placement Test consisting of 50 multiple-choice questions.

3.2.2 Reading Passage

Three intermediate text passages were selected in order to let us have a suitable reading comprehension passage for our participants based on some factors: (a) similar readability index; (b) being within the students' general knowledge and vocabulary domain; (c) having each target word once in the reading passage for the findings' reference to Laufer and Hulstijn's (2001) and not to multiple exposures; (c) Being able to extract five multiple choice questions from it in a way that finding out the meaning of the target words can be essential for answering them, that is the moderate need component of Laufer and Hulstijn's (2001) hypothesis.

Afterwards, the suggestions of the participants' teachers and intermediate students of those same institutions who were not the participants of our study were received in choosing the best text which was within our participants' readability level in terms of structure and word items. Among those reading passages, the reading passage which Walsh (2009) had used in similar investigation was chosen. His "child labor" text had 326 words with a Gunning Fog index readability of 7.73 using an on-line utility operating system. Then, the whole words which they selected as unknown for our sample population from Walsh (2009) reading passage consisted our pretest items that was administered to reveal the unfamiliar words for our participants. The pretest contained 36 items with a reliability coefficient of .62. Finally, these ten vocabulary items were selected in the text as two verbs, six nouns, one adjective, and one adverb: *plantation*, *fair*, *demonstrations*, *crops*, *sweatshop*, *fiber*, *partly*, *blame*, *march*, and *shrimp*

Giving prominence to the number of unknown words concerns with *attention* too since in this case learners can devote much of their cognitive capacity to the concept in hand (Joe, 1998). Since investigating L1 and L2 glossing in involvement load hypothesis was the purpose of our study, the reading text contained some L1 or L2 glosses in the margin.

3.3 Data Collection Procedure

Nelson Placement Test helped us to have a homogeneous sample of our selected population. Understandably, this proficiency test   let us make a more clear-cut classification based on their proficiency level. Then 22 students of the low proficient group were selected as our sample. Based on their similarities in terms of proficiency level, age, previously studied materials, and other possibly related factors, it was determined that our sample was homogeneous. Parallel with the goal of our research, they were randomly assigned to two groups in order to do two tasks.

One week before administering the tasks, a pretest was given to the participants in order to increase the reliability of our research. Since we were conducting our research in the field of incidental learning, they were not informed, in advance, of the posttests. As it was attempted not to draw the participants' attention to the lexical items, the researcher introduced the tasks as a reading activity. The researcher instructed them how to do the tasks. Since there were two different tasks and each group was required to do one of them, two different directions for each task were made clear.

The two tasks which were prepared for the participants to serve the purpose of our study were completed by each group. The participants had to do tasks with the only difference in evaluation component of task-induced involvement index. To put it simply, all tasks had moderate and strong evaluations in their constituent parts with L2 and L1 glossing, respectively. The purpose of our research was considering the effect of these tasks on enhancing the students' lexical knowledge. It was attempted that the participants not to memorize or search the target words after administering the tasks by paper collection immediately after the treatment. Besides, they did not have to communicate the target vocabularies with each other, since in accordance with Walsh (2009) it could change the need, search, and evaluation components of the hypothesis.

3.3.1 Data Collection Procedure

The selected reading text with omitted target vocabularies was prepared for Task A. It was their task to not only fill-in the gaps using ten glossed words, but also answer the comprehension questions. Each of the target words written in the margin was followed by their L2 (i.e., English) equivalents. The involvement load index of this task was two since it had a moderate need (the task required students to understand the meaning of the target vocabularies in order to fill in the blanks), no search (the students did not look the words up in the dictionary since they had been provided), and moderate evaluation (participants had to put the most appropriate words in the blanks).In order to better illustrate the constituent parts of Task A we use Hulstijn and Laufer's (2001) formulation (i.e., 1+0+1=2).

The students read similar reading passage and answered similar comprehension questions with the difference that they had to write some original sentences with target vocabularies glossed in margin in Task B. Contrary to the previous task, the participants were provided with L1 translations for the target words. According to the hypothesis, the *task-induced involvement* was calculated as three since a moderate need the (task imposed the need on the learner), no search (the meaning of the words were provided), and strong evaluation (the participants had to combine the new and the previously known words to compose an original sentence) were induced on the participants (1+0+2=3).

As time-on-task was not considered as one of the variables in the experiment, a usual time period was allotted to each task (i.e., fifteen and twenty minutes for the first and second tasks respectively). Two days later, an immediate posttest containing the ten target lexical items was given to the students and asked the participants to write their Persian or English meanings. Their test sheets were then collected and two weeks later, the same test (i.e., delayed posttest) was given to the participants with the difference that the order of its items was different. The immediate and delayed posttests were administered in five minutes.

The scoring procedure was a range between zero to one for incorrect to correct responses, respectively, which is

different from the complex method of measuring vocabulary learning in parallel experiments. Folse (2006) in his manipulation of Paribakht and Wesche's (1997) method used both receptive (asking for translations) and productive knowledge (asking for composing original sentences). Allowing students to write the words' L1 or similar equivalents to the real one can reveal if any trivial effect (receptive level) has been resulted from the experiment (Walsh, 2009). Furthermore, half a score was considered for semantically close translations of the target words which helped us in order to have a more clear-cut data collection. Consultations with other experienced teachers were done for those answers which were controversial in their semantic approximation.

3.4 Data Analysis

Nelson Placement Test gave us two homogeneous groups. Based on the normal distribution of data, using parametric or nonparametric statistical test was decided on. SPSS (version 18.0) was used for data analysis and the significance level of .05 was considered for interpreting the results.

4. Results

4.1 Testing Normality

The normality of the two hypotheses was tested to reveal the model of distribution of scores in the sample and in effect, the type of statistical tests to be conducted. The two ways of statistical and graphical tests were used to test the normality. Although each of them contains descriptive and theory driven indices; skewness, Kolmogorov-Smirnov Test, and Q-Q plot were chosen for this purpose. K-S normality index for the first experimental group revealed that at the $p < .05$ level the assumption of normality was rejected: $p = .029$. Therefore, the parametric statistic test repeated measure *ANOVA* could not be conducted and its nonparametric Friedman Test was run. On the contrary, K-S normality index for the second experimental group showed that at $p < .05$ level the assumption of normality was not rejected: $p = .782$. Since the obtained data were normal, and other assumptions of parametric techniques were met, the parametric statistic test repeated measure *ANOVA* was conducted.

4.2 Homogeneity Test Results

Based on the participants scores' in Nelson Placement Test, the homogeneity of the two groups was conducted. The K-S normality index was run to find out if the assumption of normality was met. The significant statistic of $p = .173$ at the $p < .05$ level showed that the data of the groups were normally distributed. Therefore, the parametric statistical procedure of independent-sample t-test was run. The Levene Test of equality of variances made it clear that the variances of scores across the groups were homogeneous, $p = .681$(2-tailed) at the $p < .05$. As a result, it is interpreted that there was no statistically significant difference between the means of the two groups before the treatment.

4.3 Findings of the Study

4.3.1 Testing the First Null Hypothesis

RH0 1: Using L2 glossing in involvement load hypothesis has no effect on enhancing vocabulary acquisition of Iranian EFL students.

In order to better elaborate the role of the language of glossing, L1 or L2, in task-induced involvement load, few researches can be found in the literature (e.g., Keating, 2008). The involvement load index of Task A was two since it induced moderate need (filling the gaps), no search (not using the dictionary), and moderate evaluation (putting the most appropriate words in the gaps). Therefore, we were determined to know the effect of an L2 glossed task with two involvement indices on the participants' vocabulary acquisition.

The results of the Friedman Test as can be seen in Table 2 revealed a statistically significant difference at the $p < .05$ level in the scores across the three tests (i.e., pretest, immediate posttest, and delayed posttest), χ^2 (2, $n = 11$) = 20.15, $p = .000$. The median values shown in Table 1 indicate an increase in the statistics from pretest ($Md = 2$) to immediate posttest ($Md = 6$) and a decrease in delayed posttest ($Md = 5$). These results revealed the effectiveness of L2 glossing in vocabulary acquisition but only in a short run. Seemingly, tasks with an involvement load of two with L2 glossing created a difference but that difference would be decreased by passing time.

Table 1. Descriptive Statistics of the Gap-Filling Group with L2 Gloss

	N	M	SD	Percentiles		
				25th	50th (Mdn)	75th
pretest	11	1.63	.924	1.00	2.00	2.00
immediate posttest	11	5.45	.93	5.00	6.00	6.00
delayed posttest	11	4.45	1.57	3.00	5.00	5.00

Table 2. Friedman Test for the Gap-Filling Group with L2 Gloss

N	11
Chi-square	20.15
df	2
Asymp. Sig.	.000

The nonparametric post-hoc test of Wilcoxon Signed Rank was used to pinpoint the location of the differences. Since this test required a Bonferonni adjustment to control for Type I errors, we divided the alpha level of.05 by 2 (the number of tests to be compared). Using this stricter level of .025, the results of the comparison between pretest and immediate posttest revealed a statistically significant difference, z = -3.002, p = .003, with a large effect size (d = .6), which can be seen in Table 4. Table 3 demonstrates the median scores that decreased from the pretest (Md = 2) to immediate posttest (Md = 6). According to Table 6, the comparison between immediate posttest and delayed posttest showed a statistically significant difference at the p < .025 level, z = 2.636, p = .008, with a medium effect size (d = .5) in accordance with Cohen (1988) criteria of 0.0-0.2 = small effect, 0.3-0.5 = medium effect, and 0.6-0.9 = large effect. The median values decreased from the immediate posttest (Md = 6) to delayed posttest (Md = 5) which is shown in Table 5.

Table 3. Descriptive Statistics of the Pretest and Immediate Posttest of the of the Gap-Filling Group with L2 Gloss

	N	Percentiles		
		25th	50th (Mdn)	75th
pretest	11	1.00	2.00	2.00
immediate posttest	11	5.00	6.00	6.00

Table 4. Wilcoxon Signed Rank Test for the Pretest and Immediate Posttest of the Gap-Filling Group with L2 Gloss

	immediate posttest – pretest
Z	-3.002[a]
Asymp. Sig. (2-tailed)	.003

[a] Based on negative ranks.

Table 5. Descriptive Statistics of the Immediate Posttest and Delayed Posttest of the Gap-Filling Group with L2 Gloss

	N	Percentiles		
		25th	50th (Mdn)	75th
immediate posttest	11	5.00	6.00	6.00
delayed posttest	11	3.00	5.00	5.00

Table 6. Wilcoxon Signed Rank Test for the Immediate Posttest and Delayed Posttest of the Gap-Filling Group with L2 Gloss

	delayed posttest - immediate posttest
Z	-2.636[a]
Asymp. Sig. (2-tailed)	.008

[a] Based on positive ranks.

4. 3. 2. Testing the Second Null Hypothesis

RH0 2: Using L1 glossing in involvement load hypothesis has no effect on enhancing vocabulary acquisition of Iranian EFL students.

In accordance with Laufer and Hulstijn, (2001), Task B has a moderate need (task imposed the need on the learner), no search (the meaning of the words were provided), and strong evaluation (the students had to combine the new and the previously known words to compose an original sentence) and in effect three involvement indices. The highest involvement index with L1 glossing was investigated and the differences which it would create on the vocabulary acquisition of the learners. Table 7 shows the descriptive statistics for scores of the sentence- making group with L1 glossing.

Table 7. Descriptive Statistics of the Sentence- making Group with L1 Gloss

	M	SD	N
pretest	1.18	.87	11
immediate posttest	5.72	1.48	11
delayed posttest	5.27	1.73	11

A one-way repeated measures *ANOVA* (Table 8) revealed that there was a significant difference at the $p < .05$ level among the vocabulary mean score of the three tests, Wilks' Lambda = .064, $F(2, 9) = 66.235$, $p = .000$, with a large effect size ($d = .93$) in accordance to Cohen (1988) criteria. Therefore, it is implied that the second null hypothesis was rejected. In other words, there was a statistically significant difference in vocabulary mean scores of the sentence-making group with L1 glossing who did three tests at different points of time. This finding could reveal that tasks with an involvement load of two with L1 glossing created a difference in the vocabulary acquisition of the participants. In order to pinpoint the location of the differences, the Scheffé post hoc test was conducted (Table 9) which demonstrated that each of the differences was significant.

Table 8. Repeated Measure ANOVA of Sentence- making Group with L1 Gloss

Effect		Value	F	Hypothesis df	Error df	Sig.	Cohen's d
senG1	Pillai's Trace	.936	66.235[a]	2.00	9.00	.000	.936
	Wilks' Lambda	.064	66.235[a]	2.00	9.00	.000	.936
	Hotelling's Trace	14.719	66.235[a]	2.00	9.00	.000	.936
	Roy's Largest Root	14.719	66.235[a]	2.00	9.00	.000	.936

Note. Design: Intercept Within Subjects Design: LsenG1.

[a] Exact statistic.

Table 9. Scheffé Post hoc for the Sentence- making Group with L1 Gloss

(I) LsenG1		(J) LsenG1		Mean Difference (I-J)	SE	Sig.[a]	95% Confidence Interval for Difference[a]	
							Lower Bound	Upper Bound
dimension1	1	dimension2	2	-4.545[*]	.52	.000	-6.06	-3.02
			3	-4.091[*]	.62	.000	-5.88	-2.29
	2	dimension2	1	4.545[*]	.52	.000	3.02	6.06
			3	.455[*]	.15	.049	.003	.906
	3	dimension2	1	4.091[*]	.62	.000	2.29	5.88
			2	-.455[*]	.15	.049	-.90	-.003

Note. Based on estimated marginal means.

[a] Adjustment for multiple comparisons: Bonferroni.

[*] $p < .05$ level.

5. Discussion and Conclusions

These research hypotheses examined different tasks in terms of the language of glossing in order to further investigate involvement load hypothesis. The learners' behaviors in short and long run, two days and two weeks after the treatment, were examined. In fact, these hypotheses investigated initial learning and long-term retention of the target words through the immediate and delayed posttests. It was revealed that the null hypotheses were rejected since significant differences were resulted from administering these tasks to the participants.

The first research hypothesis that considered L2 glossing made it clear that the immediate posttest was significantly superior than delayed one. Drawing on the findings of the current investigation, the claim we are led to is that Task A with L2 glossing which induces learners to compare their already known knowledge to the new ones (i.e., moderate evaluation) is superior in initial learning for the learners. However, in long-term retention we can see a poor vocabulary acquisition. On the whole, a significant decrease in the mean of the scores can be observed in administering a task with L2 glossing and two involvement indices. The observed findings were in line with the findings of incidental vocabulary

acquisition (e.g., Hulstijn & Laufer, 2001; Keating, 2008; Watanabe, 1997 ; as cited in Hui-Fang Tu, 2003). To put it simply, incidental learning can be obtained through reading or writing tasks but a significant decline can also be observed from immediate posttest to delayed posttest. In fact, this passage of time results in the inefficacy of vocabulary learning.

On the other hand, the second research hypothesis had Task B at the center of attention. To put it simply, the difference which it could create on the participants in terms of vocabulary retention was examined. Task B which had an involvement index of three made the participants compose original sentences using the target vocabulary items with L1 glossing. This research hypothesis investigated initial learning and long-term retention of the target words through the immediate and delayed posttests. A significant decline was observed between immediate and delayed posttests. Like the previous hypothesis, this poor vocabulary acquisition was consistent with the findings of incidental vocabulary acquisition. In other words, the interval between the two posttests can be considered as the main factor in the inefficacy of vocabulary learning through Task B with L1 glossing.

Although there can be found some supportive evidence in the literature (e.g., Keating, 2008; Kim, 2008; Walsh, 2009) that strengthen Laufer and Hulstijn's (2001) hypothesis, the present study did not confirm it. Therefore, at least we can claim that when there are two tasks with different language glossing, the task-induced involvement predictions can not come true. To put it simply, based on the involvement load hypothesis Task B with a higher involvement index is expected to bring about a better retention, but in these two hypotheses no sign of comparative vocabulary acquisition can be observed. In fact, it is unraveled that both tasks created similar behaviors in the participants in both tasks, that is a significant decrease in learning words. It can be concluded that the language of glossing play a part in task involvement load. Since these tasks had the only difference in the index of task-induced involvement, the evaluation component can not be regarded of so much prominence in involvement load.

Understandably, although Swain (1985 & 1995) claims that the sentence making task induces learners to devote more cognitive effort and in effect a better retention can be brought about, it seems that an L1 glossing task prevent us from such a conclusion.

5.1 Suggestions and Implications

It is implied from the findings of the current study that L2 glossed tasks with an involvement index of two and L1 glossed tasks with an involvement index of three are not effective tasks for vocabulary learning to be used by teachers or language learners.

Furthermore, even though there can be found contradictory results in the literature considering involvement load hypothesis predictions, our study found more delicate conditions in order to validate them. In other words, we can have more influential vocabulary learning if we do not trust involvement predictions when tasks have different language glossing.

5.2 Limitations of the Study

Although most of the previous empirical studies in task-induced were investigating the role of involvement load hypothesis (Laufer & Hulstijn, 2001) in incidental vocabulary learning, the present study also examined the role of the language of glossing in it. Any way, the current investigation has some limitations.

Albeit there are two arguments in the literature which are in favor of giving more control over time-on-task, we did not consider it in the present study. In the first argument, Laufer and Hulstijn (2001) state that time-on-task is an inseparable property of a task and therefore should be more finely observed. In the second argument some previous investigations (e.g., Paribakht & Wesche, 1997; Robinson, 1996, 1997; Swanborn & de Glopper, 2002; Watanabe, 1997; as cited in Hui-Fang Tu, 2003) were in favor of giving the same time of exposure to the target words in administering tasks. It is probable that we can gain the other results if we did not devote a normal time period to each task. Therefore, further investigations need to be conducted to reveal its significance in task-induced involvement with different language glossing.

Additionally, repeated exposure is a determinant factor in incidental vocabulary acquisition (Watanabe, 1997). Therefore, if the second task did not create so much repeated exposure (i.e., reading the text, answering comprehension questions, and making original sentences) on the part of the participants, we could observe Laufer and Hulstijin's (2001) claims in involvement load hypothesis. Therefore, more complementary studies are needed to give us a much clearer picture of this hypothesis especially when is it is combined with different language glossing in different tasks.

References

Cheng, H. Ch. (2011). *Vocabulary acquisition in learning English as a second language: Examining the involvement load hypothesis and language anxiety with Taiwanese college students* (Doctoral dissertation, University of Northern Colorado). Retrieved from http://www.worldcat.org/title/

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.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.

Crookes, G., & Schmidt, R. (1991). Motivation: Reopening the research agenda. *Language Learning*, *41*, 469-512.

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

Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on Foreign language vocabulary learning and forgetting. *Language Learning*, *56*(3), 463–506.

Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language and reading and vocabulary learning* (pp. 153-178). Norwood, NJ: Ablex.

Hu, H.-F., & Tu, I. R. (2003). Effects of task-induced involvement on incidental vocabulary learning in a second language. *Unpublished masters thesis, National Tsing Hua University, Taiwan*. Retrieved from http://www.hss.nthu.edu.tw/~fl/thesis/tesol/905258.pdf.

Hulstijn, J. H. (2011). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford, UK :Wiley-Blackwell.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for 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* (3), 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.

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

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, *78* (3), 285-299.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *73*, 440-464.

Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language. *Prospect*, *16*, 44–54.

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

Paribakt, T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 174-200). Cambridge University Press.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*, 27-67.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*, 45-99.

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.

Schmidt, R. (2000). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of Vocabulary Level Test. *Language Testing*, *18*(1), 55–88.

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.

Swanborn, M. S. L., & de Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning*, *52*, 95–117.

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.

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.

Won, M. (2008). *The effects of vocabulary instruction on English language learners: A meta-analysis* (Doctoral dissertation). Retrieved from http://repositories, tdle.org/ttu-ir/