The Effect of Task-induced Involvement Load on Unfamiliar L2 Vocabulary Learning: Sentence Writing, Summary Writing, Imaginary Story Writing and Creative Sentence Writing

Reyhaneh Rahmani [1], Sakineh Jafari [2*] and Siros Izadpanah [3]

[1] *MSc, Department of English, Zanjan Branch, Izlamic Azad University, Zanjan, Iran*
[2] *Assistant Professor, Department of English, Zanjan Branch, Izlamic Azad University, Zanjan, Iran*
[3] *Assistant Professor, Department of English, Zanjan Branch, Izlamic Azad University, Zanjan, Iran*

Abstract: The present study examined the effect of four types of post reading-based tasks with different index of task-induced involvement load (Laufer & Hulstijn, 2001) on EFL learners' recognition and recall of unfamiliar L2 vocabularies. To this end, 88 intermediate EFL learners were randomly assigned to four groups and were instructed to employ four different tasks after reading two narrative texts: (1) simple sentence writing; (2) text summary writing; (3) creative sentence writing; and (4) imaginary story writing. A day after the output activity session, the participants took two post-tests: the production test and the recognition test. Three weeks later, the delayed post-tests were administered. Mixed ANOVA (Split-plot) was run to compare the performances of the groups on immediate and delayed post-tests. The results revealed that there were overall significant within-group and between-group differences among four groups of the study both in immediate and delayed posttests. The creative sentence writing group outperformed in comparison to the other three groups. The results of this study turned out to be partially consistent with involvement load hypothesis.

_____

[*] **Corresponding Author.**

Authors' Email Address:
[1] **Sakineh Jafari** (sjafari591@yahoo.com), [2] **Siros Izadpana** (cyrosizadpanah@yahoo.com),

[3] **Reyhaneh Rahmani** (rahmani.reyhan@gmail.com)

Theoretical Background

Word knowledge plays a fundamental role in learning any language. As Wilkins (1972) claims "Without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111). Knowledge of vocabulary helps L2 learners in both comprehending a text and L2 development (e.g. Nation, 2001; Pulido, 2007, 2009). Second Language Acquisition (SLA) researchers have proposed a number of hypotheses to account for the processes that are involved in vocabulary learning. The involvement load hypothesis put forward by Laufer and Hulstijn (2001) proposes that the retention of unfamiliar second language (L2) word depends on the amount of involvement load of the task, i.e. the greater the learner is involved in the task the more effective the vocabulary learning. This hypothesis considers depth of processing and elaboration in learning with regard to three major task components: need, search, and evaluation. These three factors (need, search, and evaluation) are referred to as *involvement* which is regarded as a motivational-cognitive construct and predicts learners' success in retention of L2 unfamiliar words. These factors are conceptualized to vary in their degree of involvement required in performing a task: no involvement (-), moderate involvement (+), and strong involvement (++) . The 'need' factor is considered as a motivational non-cognitive dimension of involvement and is defined by Laufer and Hulstijn (2001) as "a drive to comply with the task requirements", which can be "externally imposed or self-imposed" (p. 14). On the other hand, 'search' and 'evaluation' are considered as cognitive dimensions of the involvement which are dependent upon noticing and deliberate attention to the various aspects of a word i.e. meaning or form of a word. In addition, evaluation "entails a composition 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" (p. 14). Among the three components, need has two degrees: a strong need (need ++) which is intrinsic, and a moderate need (need +) which is extrinsic. Search has only one prominence degree (search +), and entails activities such as dictionary consultation, inferencing and negotiation. The third component, evaluation, has also two varying degrees. It is moderate (evaluation +) when the decision-making process involves only comparisons, but strong (evaluation ++) when the learner creates contexts him/herself (Laufer & Hulstijn, 2001). According to Laufer and Hulstijn (2001) each of these components can be present or absent in processing a new word in different tasks.

AREL

Literature Review

A number of recent studies have submitted the involvement load hypothesis to empirical investigation and supported the idea proposed by involvement load hypothesis (e.g. Laufer & Hulstijn, 2001; Keating, 2008; Kim, 2008; Nassaji & Hu, 2012; Peters, Hulstijn, Sercu, & Lutjeharms, 2009; Rott, 2007).

A seminal study by Laufer and Hulstijn (2001), examined the rate of vocabulary learning in three types of tasks with different degrees of involvement load: reading comprehension (no involvement), reading comprehension with fill in the blank activity using target words (moderate involvement), and composition writing using target words (strong involvement). The results revealed that the composition task yielded more involvement, reading plus fill-in the blank task resulted in moderate involvement and reading comprehension led to low involvement. It was argued that the composition task which yielded the greatest retention involved a higher involvement load index compared to the other two tasks.

Similar results were also obtained by Keating (2008), who examined the effects of three types of tasks on L2 vocabulary learning: reading comprehension only, reading comprehension plus gap filling, and sentence writing. The results showed that sentence writing involved higher involvement load in learning which was in line with Laufer and Hulstijn's (2001) results.

Kim (2008) investigated whether different task-induced involvement load had any impact on the initial learning and subsequent retention of the new words. Two types of tasks were utilized: writing a composition and writing sentences, which were supposed to involve the same level of task-induced involvement load. The results showed that tasks with the same involvement loads were equally efficient for vocabulary retention. However, the researcher suggested that different degrees (i.e., moderate and strong) of the involvement load hypothesis's components (i.e., need, search, and evaluation) might not result in the same weights. Kim concluded that evaluation component might be the most efficient factor for learner's initial vocabulary learning.

Conflicting findings were obtained by Folse (2006). Three types of vocabulary tasks were used: one fill-in-the-blank activity, three fill-in-the-blanks activities, and one original sentence writing activity. According to the involvement load hypothesis, the third task was presumed to lead to the best learning; however, the results revealed that learners who performed the second task recalled the target words better than the other two groups, i.e. the task in which the learners applied each target word three times to fill in the spaces turned out to be more powerful in vocabulary retention than just sentence making exercise.

In a recent study, Rassaei (2017) studied the impacts of three kinds of activities on EFL learners' vocabulary gains. Three groups of students of English (EFL) were taught to perform three output tasks following reading texts: (1) using target vocabulary in summarizing the stories (2) answering questions with target vocabulary; and (3) predicting what would happen in the end of the story while using the target vocabulary. The results showed that predicting and answering the questions were more efficient than summarizing in learning L2 vocabulary knowledge.

In addition, Zou (2017) examined the impact of three tasks (cloze exercises, sentence-writing, and composition-writing) on learning new words by L2 learners. The results of his study was in agreement with the involvement load hypothesis: sentence and composition writing tasks with powerful involvement load resulted in better performance than cloze-exercises. Conflicting results were also found: composition-writing group was significantly more successful in recalling the words than sentence-writing group in spite of the fact that these tasks had the same involvement load index.

As the above short review recommends, SLA researchers have used different types of post-task reading activities (with different degrees of involvement load) to enhance learners' vocabulary learning. For instance, cloze exercise, text summary writing, inferring the meaning of target words, sentence completion, dictionary consultation, sentence-writing exercise, composition writing, fill in the blank exercise, and other similar kinds of tasks have been used by researchers to promote vocabulary learning of L2 learners according to the Laufer and Hulstijn's (2001) index of task-induce involvement load.

Although previous researches have showed that written output would be effective for the acquisition of L2 vocabulary, more studies are needed to shed further light on the effectiveness of the various writing activities on the acquisition of L2 word learning. For example, written output activities can be distinguished in terms of whether they require creative writing such as writing an imaginary story or if they entail restating or reconstructing previously obtained information such as summarizing a text (Rassaei, 2017). It is therefore interesting to know which one would be more conducive to L2 vocabulary learning.

Therefore, the current study made an attempt to investigate the effects of four reading-based activities on the acquisition of unfamiliar L2 vocabularies. The four writing activities which were examined in the current study were: (1) writing a sentence for each new word in the text (2) summarizing a story by using the new words (3) writing a creative sentence for each new word, and (4) writing an imaginary story by using the new words. In order to fulfill the purpose of the present study the following research questions were formulated:

Q1: Do types of reading based tasks (sentence writing, different stories, summarizing text, and imaginary sentence writing) have any significant effect on the retention and recall of the unfamiliar L2 vocabulary of the Iranian EFL learners probed through an immediate and delayed multiple choice recognition test (MCRT)?

Q2: Do types of reading based tasks (sentence writing, different stories, summarizing text and imaginary sentence writing) have any significant effect on the retention and recall of the unfamiliar L2 vocabulary of the Iranian EFL learners probed through an immediate and delayed cued response production test (CRPT)?

Method

Participants

In order to provide answer to the research questions, 88 intermediate EFL learners studying English at a private language institute in Zanjan, Iran enrolled in this study. The participants were chosen among the learners who had been placed at the intermediate level based on the institute's placement test. However, to assure the homogeneity of the participants for the present study, they were selected (42 males and 46 females) out of 110 leaners according to the results of Nelson placement test. Based on the mean scores, students whose score fell one standard deviation below and above the mean, 88 students were selected. Then, they were randomly assigned to four groups with 22 participants in each group. They ranged in age from 17 to 27 year and had an average English learning experience of 3 years in EFL classes. While the data were gathered, the participants were attending English classes three times a week. Following their consent to participate in the study, they were randomly assigned into four groups (22 learners in each group) to examine the effect of different types of written reading-based tasks on their retention of unfamiliar L2 vocabularies. Table 1 shows the number, gender, and the grouping of the participants.

Table 1. Grouping and Demographic Information of the Participants

	Groups	*Numbers*	*Age*	*Gender*
1	Sentence writing	22	17-27	12 males 10 females
2	Different story writing	22	17-27	10 males 12 females
3	Summary writing	22	17-27	11 males 11 females
4	Creative sentence writing	22	17-27	9 males 13 females
Total	**4**	**88**	**17-27**	**42 males 46 females**

Materials

The instruments utilized in the present study were as follows:

Testing Materials

1. Pretest

In order to identify the target words for the present study (words which were new and unfamiliar to the participants) two reading passages were selected from chicken Soup for the Soul (Canfield & Hansen, 1993). Great care was exercised to choose passages which are not so long or complicated. Also, an attempt was made to select interesting texts which would capture learners' attention. Based on Rassaei (2017), from these two passages 40 vocabulary items were selected by the researchers because they were unlikely to be familiar to intermediate learners. As a pre-test and in order to identify target vocabulary items, the learners were required to answer the following two questions for each of these 40 selected vocabulary items: (1) whether or not the word is familiar to them; and, (2) if the word is familiar, what is its first language (L1) translation or English definition? Based on the result of the pretest (on 40 words) 14 target words which were unfamiliar to the learners were chosen (7 words in each reading passage). A range of parts of speech was covered in the selection of the target words in order to eliminate any possible effect of grammatical category, i.e. noun, verb, adjective, and adverb on vocabulary learning (Zou, 2017).

2. Posttests

Following Rassaei (2017), two tests were prepared and administered as immediate and delayed post-tests in order to assess the participants' L2 vocabulary retention and recall. They were a cued response production test *(CRPT)* and a multiple choice recognition test (MCRT). No time limitation was set for taking the tests. For *Cued response production test:* each item included one sentence (or a combination of two sentences) in which target vocabulary item was simply replaced with an L1 translation along with the first English letter of the target word. The number of letters for each target word was also specified. The participants were asked to read the sentences and fill in the missing letters. For scoring the test, spelling errors were ignored as long as the word was understandable to the rater. For example, for the target word "fringe" the following item was written:

Example 1: There is a beautiful view of **f - - - - -** (حاشیه) of trees behind the river banks.

To estimate the reliability of the test, test-retest method was applied. The correlation coefficient between the scores of a group of learners (which were not participated in the main study) in two performances of the test was calculated which established to be .86.

For Multiple Choice Recognition Test (MCRT) 14 target vocabulary items were designed. Each item included a statement in which a word was removed and followed by four options. The participants were asked to choose one of the four choices in order to complete the statement. An example of the MCRT is provided in Example 2:

Example 2:

14 – We had to …………. the meeting when our boss felt ill.

A: adjourn

B: advance

C: start

D: convene

Again, to estimate the reliability of the test, test-retest method was applied. The correlation coefficient between the scores in two performances of a group of learners (which were not participated in the main study) was calculated which established to be 0.83.

Reading Passages

Ten short passages were selected from chicken Soup for the Soul (Canfield, Hansen, & Unkovich, 2007). These passages were given to 4 experts in order to select two short narratives. Great care was exercised to choose passages which are not so long or complicated. Also, an attempt was made to select interesting texts which would capture learners' attention. Two narrative passages were selected by the experts to serve the purposes of the study.

In the output activity session, the learners were exposed to both passages. The passages included 14 target vocabulary items (each passage 7 words). The target words that were underlined were followed by their first language (L1) definition embraced in parentheses. The passages were the same for the four experimental conditions. Each passage included 600 words approximately.

Vocabulary Learning Tasks

To achieve the purpose of the study four post reading-based vocabulary tasks were assigned to four groups of participants:

1. Sentence writing task:

In this type of activity learners were asked to write an original sentence using the target vocabulary. According to Laufer and Hulstijn's (2001) hypothesis we categorized the involvement components of this task as (need +, search -, evaluation ++) since the need was

externally imposed, word meaning (L1 translation) was provided, and creation of original contexts was involved.

2. Summary writing task:

In this type of activity learners were asked to write a summary of the reading text and incorporate the key vocabularies, introduced in the reading, in their summary writing. Correct use of all words were required for task completion. According to Laufer and Hulstijn's (2001) hypothesis we categorized the involvement components of this task as (need +, search -, evaluation +) since the need was externally imposed, word meaning (L1 translation) was provided, and reproduction of the same context was involved.

3. Creative sentence writing task:

In this type of activity learners were asked to write an original sentence using the target vocabulary by utilizing alliteration (e.g. V̲ahid **v̲anished** with his v̲an when his mum asked him to clean the house with a v̲acuum cleaner).

Alliteration is defined by Eyckmans and Lindstromberg *(2016)* as the occurrence of the same consonant onset in two or more content words within a phrase (e.g. *miss the mark*). Research findings show that when participants are asked to recall word pairs in whatever order they like, pairs that show sound repetition are learned faster and recalled better than ones that do not (e.g. Bower & Bolton, 1969; Lindstromberg & Boers, 2008; Eyckmans & Lindstromberg, *2016*). Based on Laufer and Hulstijn's (2001) hypothesis we categorized the involvement components of this task as (need +, search -, evaluation ++) since the need was externally imposed, word meanings (L1 translation) were provided, and the creation of original context was involved.

4. Imaginary Story writing task:

In this type of activity learners were asked to write an imaginary story and incorporate the target vocabularies, introduced in the readings, in their story writing. Correct use of all words were required for task completion. Based on Laufer and Hulstijn's (2001) hypothesis we categorized the involvement components of this task as (need +, search -, evaluation ++) since the need was externally imposed, word meanings (L1 translation) were provided, and the creation of original context was involved.

Procedure

The participants were randomly put into four groups. At the beginning of the study, a pre-test was administered to select vocabulary items that were unfamiliar to learners and would thus

serve as target vocabulary in this study. One week later, treatment session began for four groups (before treatment each group received the required instruction for completion of the post-reading task). A day after the output activity session, the participants took the unannounced immediate post-tests (MCRT and CRPT). Then after three weeks, the unannounced delayed post-tests (MCRT and CRPT) were given to them.

Data Analysis

To analyze the obtained data, Mixed ANOVA (Split-plot) was run through SPSS Software Version 23 to compare the performance of the groups on vocabulary immediate and delayed post-tests. The reason behind choosing Mixed ANOVA for analyzing the present study's data was that this test compares the mean difference between groups that have been split on two "factors" (known as independent variables), where one factor is a 'between subjects' factor (here in this study the type of post-reading vocabulary activities) and the other factor is 'within-subjects' factor (the time interval of the participants' performance from immediate posttest to delayed posttest).

Results of Testing the Normality Assumption

Before analyzing the obtained data the normality of the distribution of the groups' scores was ascertained by using Kolmogorov-Smirnov test (K S- test). The results of K S-test (Table 2) indicated a normal distribution of the scores for the groups since the p value exceeded .05.

Table 2. *Kolmogorov-Smirnov Test for Checking the Normality of the Distribution of Scores*

	Kolmogorov-Smirnov Z	*P value*	*Result*
Writing sentence group	1.513	.265	Normal
Writing summary group	.987	.489	Normal
Writing imaginary story group	1.645	. 542	Normal
Writing creative sentence group	.765	.345	Normal

After ensuring the normality of the distribution of data, to examine the statistical significance of the difference in the mean scores of the groups, Mixed ANOVA was run.

Results of the Mixed ANOVA for MCRT Test

To provide answer to the research questions 1, the obtained data were analyzed using SPSS software version 23. The results of the mean scores of the four groups in Multiple Choice Recognition Test (MCRT) on immediate and delayed posttests are presented in Table 3 and Table 4 respectively.

Table 3. *Descriptive Statistics; Immediate Posttest of Multiple Choice Recognition Test by Groups*

Groups	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Creative Sentence	22	9.59	3.246	.692	8.15	11.03
Different Story	22	8.73	2.848	.607	7.46	9.99
Summarizing Text	22	8.59	2.823	.602	7.34	9.84
Sentence Writing	22	6.09	3.235	.690	4.66	7.53
Total	**88**	**8.25**	**3.267**	**.348**	**7.56**	**8.94**

Table 4. *Descriptive Statistics; Delayed Posttest of Multiple Choice Recognition Test by Groups*

Groups	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Creative Sentence	22	8.14	3.091	.659	6.77	9.51
Different Story	22	7.86	2.455	.523	6.78	8.95
Summarizing Text	22	6.77	2.202	.470	5.80	7.75
Sentence Writing	22	5.05	2.591	.552	3.90	6.19
Total	**88**	**6.95**	**2.836**	**.302**	**6.35**	**7.56**

To statistically compare the four groups in terms of 'within-subject' and 'between-subject' differences in MCRTs a Mixed ANOVA was run. Table 5 shows the results of Mixed-design ANOVA for overall significant differences in MCRT test scores (immediate and delayed posttests).

Table 5. *The Results of Mixed-design ANOVA for Overall Within-group and Between-group Significant Differences among Four Groups of the Study in MCRT Test*

		SS	df	MS	F	Sig.	ηp^2
Within-group							
	Time	73.841	1	73.841	9.511	.003	.102
	Group*Time	6.023	3	2.008	.259	.859	.776
	Error	652.136	84	7.766			
Between-group							
	Group	273.477	3	91.159	10.991	.000	.282
	Error	696.682	84	294.798			

As depicted in the table, there were overall significant within-group and between-group differences (P-values $\leq$ .05) among four groups of the study. In addition, no interaction was found between two factors of group and time. This means that the interval time between immediate and delayed posttests did not have any significant effect on different groups' performance. Although learners got lower scores in immediate posttest compared to delayed posttest, this poor performance was detected for all of the groups. Yet, in order to determine

significant pairwise within-group differences and to be able to answer research question 1 Bonferroni adjusted multiple comparisons were conducted in the Mixed-design ANOVA procedure (Table 6). The results showed that unlike group 3 with a significant within-group difference between the posttest and delayed posttest scores (p-value = .033), there were no significant differences in the posttest and delayed posttest scores of other three groups (group1: p-value = .087; group 2: p-value = .307; group 4: p-value = .217). The group differences in MCRT test, in terms of interval time between immediate and delayed posttests, were displayed in profile plot in figure 1 (see Appendix 1). It displays the profile plot of MCRT in terms of interval time between immediate and delayed posttests in different groups. As it is shown, imaginary sentence writing group (group 1) has the highest mean scores compared to the other groups both in immediate and delayed posttests. The figure also shows that imaginary story writing group (group 2) recognized the words in the delayed posttest better than the other three groups, i.e. the performance of group 2 did not change significantly in terms of the time interval between immediate and delayed posttests of MCRT.

Therefore, the null hypotheses 1 can be rejected according to the above mentioned results.

Table 6. *The Results of Bonferroni Adjusted Multiple Comparisons for Significant Between-group Differences for MCRT (group 1= creative sentence, group 2=imaginary story, group 3= summarizing text, group 4= sentence writing)*

Test	Groups	p-value
	group 1/ group 2	1.000
	group 1/ group 3	1.000
	group 1/ group 4	.002
Immediate Posttest	group 2/ group 3	1.000
	group 2/ group 4	.031
	group 3/ group 4	.047
	group 1/ group 2	1.000
	group 1/ group 3	.517
	group 1/ group 4	.001
Delayed posttest	group 2/ group 3	1.000
	group 2/ group 4	.003
	group 3/ group 4	.184

As shown in Table 6, there were significant differences between groups 1 and 4, groups 2 and 4, and groups 3 and 4 in the immediate posttest scores. There are also significant between-group differences between groups 1 and 4 and groups 2 and 4 in the delayed posttest.

Results of Mixed ANOVA for CPRT Test

To find answer to the research questions 2 the obtained data were analyzed through SPSS software. The results of mean scores of the four groups in Cued Response Recognition Test (CPRT) on immediate and delayed posttests are presented in tables 7 and 8, respectively.

Table 7. *Descriptive Statistics; Posttest of Cued Response Test by Groups*

Groups	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	*Upper Bound*
Creative Sentence	22	9.41	2.789	.595	8.17	10.65
Different Story	22	9.23	2.776	.592	8.00	10.46
Summarizing Text	22	7.73	2.763	.589	6.50	8.95
Sentence Writing	22	5.32	3.315	.707	3.85	6.79
Total	**88**	**7.92**	**3.309**	**.353**	**7.22**	**8.62**

Table 8. *Descriptive Statistics; Delayed Posttest of Cued Response Test by Groups*

Groups	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	*Upper Bound*
Creative Sentence	22	7.95	2.400	.512	6.89	9.02
Different Story	22	7.45	2.558	.545	6.32	8.59
Summarizing Text	22	6.50	2.283	.487	5.49	7.51
Sentence Writing	22	5.09	2.524	.538	3.97	6.21
Total	**88**	**6.75**	**2.640**	**.281**	**6.19**	**7.31**

To statistically compare the four groups in terms of 'within-subject' and 'between-subject' differences in CRPTs a Mixed ANOVA was run. Table 9 shows the results of Mixed-design ANOVA for overall significant differences in CPRT test scores.

Table 9. *The results of Mixed-design ANOVA for overall within-group and between-group significant differences among four groups of the study in CPRT test*

		SS	df	MS	F	Sig.	ηp^2
Within-group							
	Time	60.278	1	60.278	7.844	.006	.085
	Group*Time	14.699	3	4.900	.638	.593	.022
	Error	645.523	84	7.685			
Between-group							
	Group	326.199	3	108.733	15.953	.000	.363
	Error	696.682	84	294.798			

As depicted in Table 9, there were overall significant within-group and between-group differences (*P*-values ≤ .05) among four groups of the study In addition, no interaction was found between two factors of group and time. This means that the interval time between immediate and delayed posttests did not have a significant effect on different groups'

performance. Although learners got lower scores in immediate posttest compared to delayed posttest, this poor performance was detected for all of the groups. Nevertheless, in order to determine significant pairwise within-group differences and to answer research question 2, Bonferroni adjusted multiple comparisons were conducted in the very Mixed-design ANOVA procedure (Table 10). The results showed that unlike group 2 with a significant difference between the posttest and delayed posttest scores (p-value = .037), there were no significant differences in the posttest and delayed posttest scores of other three groups (group 1: p-value = .085; group 3: p-value = .146; group 4: p-value = .786). The group differences in CRPT test, in terms of interval time between immediate and delayed posttests, were displayed in profile plot in figure 2 (see Appendix 2). This figure displays the profile plot of CRPT in terms of interval time between immediate and delayed posttests in different groups. As it is shown, imaginary sentence writing group (group 1) has the highest mean scores compared to the other groups both in immediate and delayed posttests. It also shows that the sentence writing group (group 4) recalled the words in the delayed posttest better than the other three groups, i.e. the performance of group 4 did not change significantly in terms of the time interval between immediate and delayed posttests of CRPT.

Table 10. *The Results of Bonferroni Adjusted Multiple Comparisons for Significant Between-group Differences in CRPT (group 1= creative sentence, group 2=imaginary story, group 3= summarizing text, group 4= sentence writing)*

Test	Groups	p-value
Immediate Posttest	group 1/ group 2	1.000
	group 1/ group 3	.357
	group 1/ group 4	.000
	group 2/ group 3	.553
	group 2/ group 4	.000
	group 3/ group 4	.046
Delayed posttest	group 1/ group 2	1.000
	group 1/ group 3	.310
	group 1/ group 4	.001
	group 2/ group 3	1.000
	group 2/ group 4	.011
	group 3/ group 4	.356

As shown in Table 10, there were significant differences between-group 1 and 4, groups 2 and 4, and groups 3 and 4 in the immediate posttest scores. There were also significant between-group differences between groups 1 and 4 and groups 2 and 4 in the delayed posttest.

Therefore, we can conclude that writing a sentence, writing an imaginary story, writing an imaginary sentence, and writing a summary of the story for each of the target vocabulary have an effect on EFL learners' subsequent retention and recall of those target words. The Mixed ANOVAs results showed that the imaginary sentence writing group outperformed in comparison to the other three groups, both in immediate (MCRT & CRPT) and delayed (MCRT & CRPT) posttests. The imaginary story writing, summary story writing and sentence writing groups were at the second, third, and fourth places respectively (research question 1 and 2). The null hypotheses 1 and 2 can be rejected accordingly.

Discussion

The involvement load hypothesis put forward by Laufer and Hulstijn (2001) predicts that the retention of an unfamiliar second language (L2) word depends on the amount of involvement load of the task, that is, the greater the learner is involved in the task the more effective the learning. Based on the Laufer and Hulstijn's (2001) index of task-induced involvement load, the present study examined the effect of four different types of post reading-based tasks on EFL learners' vocabulary learning. To this end, after reading two narrative texts, four groups of intermediate EFL learners were instructed to employ four different reading-based tasks: (1) simple sentence writing; (2) text summary writing; (3) creative sentence writing; and (4) imaginary story writing.

To answer the research questions of the study, two Mixed ANOVAs were run. The results revealed an overall significant within-group and between-group differences among four groups of the study both in immediate and delayed posttests (MCRT and CRPT). The creative sentence writing group outperformed the other three groups, both in immediate (MCRT & CRPT) and delayed posttests (MCRT & CRPT). The imaginary story writing, summary story writing and sentence writing groups were at the second, third, and fourth places respectively (creative sentence writing > imaginary story writing > summary writing > sentence writing) . Therefore, regarding the first research question (Do types of reading based tasks have any significant effect on the retention and recall of the unfamiliar L2 vocabulary of the Iranian EFL learners probed through an immediate and delayed multiple choice recognition test?), and the second research question (Do types of reading based tasks have any significant effect on the retention and recall of the unfamiliar L2 vocabulary of the Iranian EFL learners probed through an immediate and delayed cued response production test?) our results suggest that the answers to these two questions are 'yes'.

This general finding is in agreement with the Laufer and Hulstijn's (2001) index of task-induced involvement load, since these four reading based tasks provided learners with unequal amount of involvement load while they processed the target words in performing the tasks. As Laufer and Hulstijn's (2001) hypothesis suggests output tasks will not necessarily lead to better results than the input tasks, (i.e. what is important is the degree of motivational-cognitive dimension or involvement load which is offered by the task). The findings of recent similar studies (e.g. Hu & Nassaji, 2016; Keating, 2008; Rassaei, 2017; Zou, 2017) also confirm Laufer and Hulstijin's proposal and are in line with the present study's findings.

Moreover, one can argue that the best performance which was achieved by creative writing group (and imaginary story writing group) may be the result of longer time spent by learners on the task. However, this longer time-on-task can be considered as the byproduct of the cognitive processes induced by the type of task rather than other factors (Rassaei, 2017). As Laufer and Hulstijn (2001) indicated the time required to complete a task can be regarded as an inherent characteristic of the task which simply imposes more cognitive load and processing on the part of the learners. As the findings of the existing studies (e.g. Chen, 2002; Erten & Tekin, 2008; Hill & Laufer, 2003; Nakata, 2008; Hu & Nassaji, 2016) on the relationship between time-on-task and task effectiveness show, spending much time on a task does not necessarily guarantee the subsequent retention of the new words. According to Hu and Nassaji (2016), the time on task factor did not have any impact on the overall findings of their study since the higher involvement load tasks that entailed longer time to be completed did not necessarily result in better performance of the learners.

Conflicting findings with involvement load hypothesis also emerged. The findings of this study did not support the proposal formulated by the Involvement Load Hypothesis which assumes that the tasks with equal involvement load would result in similar learning gains. Although the hypothesis specifies sentence writing, creative sentence writing, and imaginary text writing as inducing the same load (in terms of need+, search-, and evaluation++), the creative sentence writing group outperformed the other three groups. This finding, as proposed by other researchers (e.g. Hu & Nassaji, 2016; Rassaei, 2017; Zou, 2017), addresses the need for refining or expanding the predictive power of the index of task-induced involvement load (need, search, and evaluation) in order to make it possible to categorize the various types of vocabulary learning tasks in a more efficient way. As proposed by Zou (2017) an augmented evaluation framework is needed to categorize some tasks in terms of evaluation induced by various tasks as 'moderate +' evaluation, 'strong ++'

evaluation, and 'very strong +++' evaluation. In this manner, the evaluation induced by creative sentence writing task triggered greater task involvement compared to the other three tasks and can be categorized as imposing a 'very strong +++' evaluation. In addition, this can be explained readily in terms of a more elaborated and highly creative processing required by the task in incorporating alliterative words in the production of the sentences since the creative sentence writing group not only required to produce a new sentence for the target word (like sentence writing group) but also required to use as many alliterative words as possible in the sentence. This finding is in line with recent experimental studies (e.g. Bores, Lindstromberg, & Eckmans, 2008; Eckmans & Lindstromberg, 2016; Lindstromberg & Boers, 2008; Lindstromberg & Eckmans, 2014) which have found that alliterative word combinations (first sound/consonant repetition in a string of two or more words within a phrase e.g. make a mistake) make these word combinations more memorable to the language learners.

The second task which turned out to be more effective in participants' subsequent word recalls was imaginary story writing task. This finding is consistent with previous research results in which creating a new sequence of events proved to be more demanding for learners than using the target words in a familiar sequence of events (e.g. Joe, 1998; Rassaei, 2017; Zou, 2017) as this task triggered deep involvement than the summary writing task. This task which asked participants to incorporate target words in a new context was considered by Laufer and Hulstijn (2001) as carrying out a strong evaluation (++) component since in this task "evaluation requires making a decision about additional words which will combine with the new word in an original sentence or text" (p. 15). The results of a study by Joe (1998) showed that new words when were used in a learner-generated original text would be retained better compared to the words which were used in a non-original text, since according to Laufer and Hulstijn (2001), these words were "underwent a higher involvement load than words which were retained less well" (p. 20).

Another finding of this study, which appeared to be inconsistent with the involvement load hypothesis, was that the summary writing group gained better scores in posttests compared to the sentence writing group. Although task-induced involvement load of sentence writing is hypothesized to be more than summary writing task (since according to Laufer and Hulstijn's hypothesis the involvement load components of summary writing task is categorized as (need +, search -, evaluation +) as the need was externally imposed, word meaning (L1 translation) was provided, and reproduction of the same context was involved,

and the sentence writing task is categorized as (need +, search -, evaluation ++) since the need was externally imposed, word meaning (L1 translation) was provided, and creation of original contexts was involved) the summary writing group outperformed the sentence writing group. In order to account for this finding, one can argue that producing a text may lead to more cognitive processing and seems to be more demanding than writing the isolated sentences. This would be due to the different mental processing required by each task. Another possible explanation would be that because participants were informed in advance to prepare a summary of the story after reading, they put more effort in analyzing the text, identifying the important information, and making association between the target words and the gist of the text. The results of the studies (e.g. Marzec-Stawiarska, 2016; Oded & Walters, 2001) on the relationship between reading and summarizing reveal that summarizing engages learners in an in-depth analysis of the text which results in better performance of different post reading-based tasks.

Conclusion

Based on the arguments made by the previous studies (e.g. Jahangard, 2013; Keating, 2008; Rassaei, 2017; Yaqubi, Rayati & Allemzade Gorgi, 2010; Zou, 2017) we came to the conclusion that there is a need for refining or expanding the predictive power of the indexes of task-induced involvement load (need, search, and evaluation) in order to make it possible to categorize the various types of vocabulary learning tasks in a more efficient way. The findings of this study did not support the proposal formulated by the Involvement Load Hypothesis which assumes that the tasks with equal involvement load would result in similar learning gains. Although the hypothesis specifies sentence writing, creative sentence writing, and imaginary text writing as inducing the same load (in terms of need+, search-, and evaluation++), the creative sentence writing group outperformed than the other three groups. Therefore, a detailed evaluation index is needed in order to categorize the evaluation induced by various tasks as 'moderate +' evaluation, 'strong ++' evaluation, and 'very strong +++' evaluation. In this manner, the evaluation induced by creative sentence writing task triggered greater task involvement compared to the other three tasks and can be categorized as imposing a 'very strong +++' evaluation. Thus, as the findings suggest more elaborated and highly creative processing required by the task increases its evaluation degree.

Implications of the Study

The results of the present study have considerable practical implications for language pedagogy in various EFL contexts. First, as the findings of this study suggest vocabulary gains of learners would be affected as a result of utilizing different post-reading vocabulary learning tasks. Furthermore, the results indicate that among the writing tasks, those that engaged participants' creativity and imagination led to an efficient vocabulary retention. Acknowledging the benefits of creative writing, Nunan (1999) also suggests that "the use of imaginative literature could be used much more extensively than it is" (p. 76).

The second implication of this study is that teachers must be more sensitive to the tasks they use in their classrooms. Engaging and challenging tasks may generate the necessary motivation in L2 learners in acquiring new words in an efficient manner.

Limitations of the Study and Suggestions for Future Research

Despite the above mentioned points, it should be indicated that the present study is not without its limitations. First, limited subjects from a single language institute were participated in this study. Further research is also needed in different contexts to compare the results. Second, only four types of vocabulary learning tasks were examined in this study. Future research will be undertaken to investigate the effect of other tasks, with similar or different degrees of involvement load index on EFL learners' vocabulary learning. In addition, time on task was not controlled in this study and this could be considered as a limitation. Future studies can attempt to focus on this limitation and investigate the effect of this variable in their studies. Also, in this study, participants were informed in advance about what output task would be followed by the reading input. This would affected the processes they had undertaken in reading the texts. Therefore, future research can be conducted to examine whether different results can be achieved if the participant are not informed in advance about a particular output activity after reading the text.

References

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). When does assonance make lexical phrases memorable? *European Journal of Applied Linguistics and TEFL*, 3, 93–107.

Bower, G., & Bolton, L. (1969). Why are rhymes easy to learn? *Journal of Experimental Psychology,* 82, 453–461.

Canfield, J. & Hansen, M. & Unkovich, J. (2007). *Chicken soup for the soul*. Random House: UK Limited.

Chen, H. (2002). *Investigating the effects of L1 and L2 glosses on foreign language reading com-prehension and vocabulary retention*. Unpublished paper presented at the annual meeting of the Computer-Assisted Language Instruction Consortium, Davis, CA, USA.

Erten, H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, 36, 407–422.

Eyckmans, J. & Lindstromberg, S. (2016). The power of sound in L2 idiom learning. *Language Teaching Research,* 4, 1-21.

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

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics*, 41, 87–106.

Hu, H. & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement Load Hypothesis versus Technique Feature Analysis. *System*, 59, 28-39.

Jahangard, A. (2013). Task-induced involvement in L2 vocabulary learning: a case of listening comprehension. *Journal of English Language Teaching and learning,* 12, 43-62.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19, 357–377.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research,* 12, 365– 386.

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

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

Lindstromberg, S., & Boers, F. (2008). Phonemic repetition and the learning of lexical Chunks: The mnemonic power of assonance. *System*, 36, 423–436.

Lindstromberg, S., & Eyckmans, J. (2014). How big is the positive effect of assonance on the near-term recall of L2 collocations*? International Journal of Applied Linguistics*, 165, 19–45.

Marzec-Stawiarska, M. (2016). The influence of summary writing on the development of reading skills in a foreign language. *System*, 59, 90-99.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *Re CALL*, 20, 3–20.

Nassaji, H., & Hu, M. (2012). The relationship between task-induced involvement load and learning words from context. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 5, 69-86.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

Nunan, D. (1999). *Second Language Teaching and Learning*. Boston, MA: Heinle & Heinle.

Oded, B., & Walters, J. (2001). Deeper processing for better EFL reading comprehension. System, 29,357-370.

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: the effect of three enhancement techniques compared. *Language Learning*, 59 (1), 113-151.

Pulido, D. (2007). The effect of topic familiarity and passage sight vocabulary on L2 Lexical inferencing and retention through reading. *Applied Linguistics*, 28(1), 66-86.

Pulido, D. (2009). How involved are American L2 Learners of Spanish in Lexical input processing tasks during reading? *Studies in second language Acquisition*, 31, 31-58.

Rassaei, E. (2017). Effects of three forms of reading – based output activity on L2 vocabulary learning. *Language Teaching Research*, 3, 1-20.

Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, 57 (2), 165-199.

Wilkins, D. A. (1972). *Linguistics in Language Teaching. London*: Edward Arnold.

Yaqubi, B, Rayati, R., & Allemzade Gorji, N. (2010). The involvement load hypothesis and vocabulary learning: the effect of task types and involvement index on L2 vocabulary acquisition .*The Journal of Teaching Language Skills*, 2(1), 145-163.

Zou, D. (2017). Vocabulary acquisition through close exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research,* 4, 1-22.
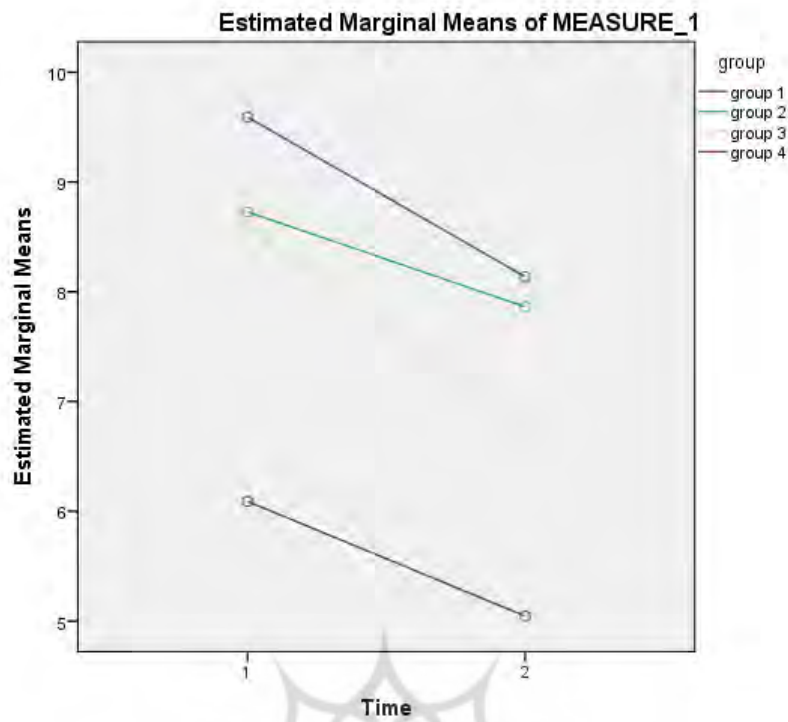
Appendix 1

Figure 1. Profile Plot *for MCRT (group 1= imaginary sentence, group 2=imaginary story, group 3= summarizing text, group 4= sentence writing)*
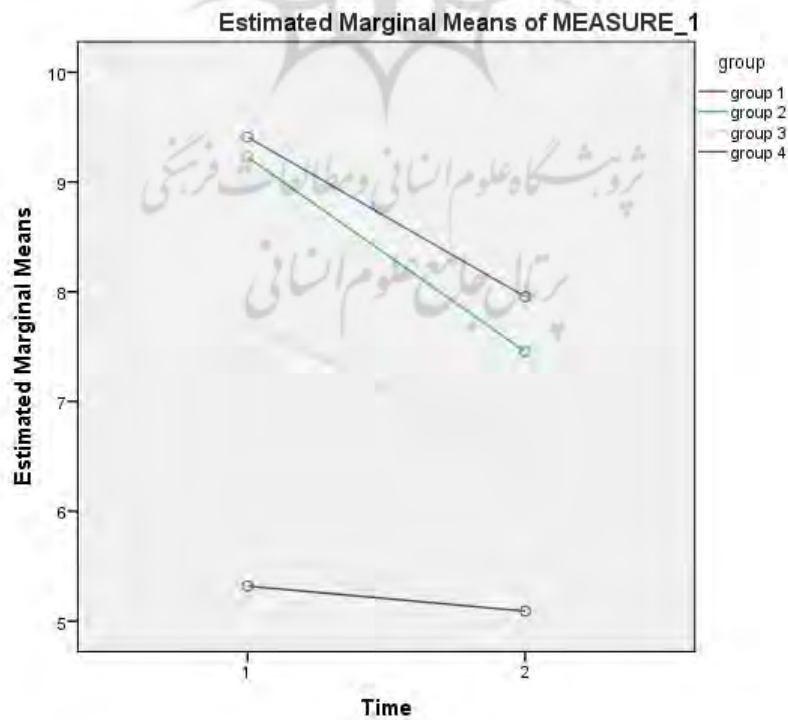
Appendix 2

Figure 2. Profile Plot of CRPT *(group 1= creative sentence, group 2=imaginary story, group 3= summarizing text, group 4= sentence writing)*