Predictive Power of Involvement Load Hypothesis and Technique Feature Analysis across L2 Vocabulary Learning Tasks

Nahid Chaharlang, Department of English Language Teaching, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
Department of English Language Teaching, Khouzestan Science and Research Branch, Islamic Azad University, Ahvaz, Iran
chaharlangnahid@gmail.com
Mohammad Taghi Farvardin*, Department of English Language Teaching, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
farvardin@iauahvaz.ac.ir

Abstract

Involvement Load Hypothesis (ILH) and Technique Feature Analysis (TFA) are two frameworks which operationalize depth of processing of a vocabulary learning task. However, there is dearth of research comparing the predictive power of the ILH and the TFA across second language (L2) vocabulary learning tasks. The present study, therefore, aimed to examine this issue across four vocabulary learning tasks (i.e., reading with glosses, keyword techniques, word card, and reading and finding the words in text) ranked differently by the ILH and the TFA. To this end, 80 English as a foreign language (EFL) learners were randomly assigned to one of four tasks of learning 16 target words. The results of one-way ANOVA, LSD Post hoc tests, and multiple regression analyses showed that the TFA had a better explanatory power than the ILH in predicting vocabulary learning gains. The findings highlight the TFA as a more powerful framework.

Keywords: Involvement Load Hypothesis, Technique Feature Analysis, Vocabulary Learning Task, Vocabulary Learning, EFL Learners

Introduction

Vocabulary has been regarded as the building block of second language (L2) learning (Knight, 1994; Nation & Webb, 2011). Therefore, it is necessary for language teachers to expand and arrange well-chosen and applicable vocabulary learning tasks (Bao, 2015). However, it is important to know which vocabulary task can be more effective (Nation & Webb, 2011).

Hulstijn and Laufer (2001) claimed that retention of unfamiliar words is attributable to the amount of involvement while processing the words. In the same line, Hulstijn and Laufer (2001) proposed the Involvement Load Hypothesis (ILH). According to Nation and Webb (2011), the ILH is "the best-known and best-researched way of analyzing vocabulary teaching techniques" (p. 3). In the ILH, involvement load (IL) is regarded as a motivational-cognitive construct which has three elements including *need*, *search*, and *evaluation* which conceptualize the depth of processing (Laufer & Hulstijn, 2001). *Need* refers to whether learning an unknown word is needed or not; *search* refers to whether the learner is forced to explore or restore the meaning or form of a specific word; *evaluation* indicates whether a task requires learners to relate the form or meaning of an unknown word with different words in the context. Each of these elements may be absent (-), present with moderate strength (+), or present with full strength (++). Hulstijn and Laufer (2001) suggested that "absence of a factor is marked as 0, a moderate presence of a factor as 1, and strong presence as 2" (p. 544). As a result, zero to six IL indices can be given to a task. The total strength of these three factors shows the IL of the task. In other words, more IL would

lead to better learning (Nation & Webb, 2011). The research on the ILH has encouraged scholars to develop other frameworks to predict the effectiveness of L2 vocabulary learning tasks.

As a complementary framework for the ILH, Nation and Webb (2011) introduced Technique Feature Analysis (TFA) representing some other criteria for the operationalization of depth of processing. This framework involves five components (i.e., motivation, noticing, retrieval, generation and retention). Hence, the TFA framework involves 5 components which propose 18 criteria to evaluate vocabulary learning tasks (Nation & Webb, 2011). As shown in Table 1, *motivation* deals with a clear learning goal for a vocabulary learning activity. *Noticing* concerns the attention an activity gives to the target words and whether it encompasses negotiation which happens when learners intentionally study a word, guess from context, or have a word explained to them (Nation, 2001). *Retrieval* includes recall of words and whether there are multiple retrievals between each interval. *Generation* can be either receptive or productive (Nation, 2001). Receptive generation refers to facing a word while listening to or reading an unfamiliar context, while productive generation is using the word in new contexts. Finally, *retention* deals with successful association of form and meaning. That is, whether it encompasses instantiation, imaging, and avoids interference.

Table 1. *Technique Feature Analysis (adopted from Nation & Webb, 2011, p. 7)*

Criteria	Scores (0-1)
Motivation	
Is there a clear vocabulary learning goal?	
Dose the activity motivate learning?	
Do the learners select the words?	
Noticing	
Dose the activity focus attention on the target words?	
Dose the activity raise awareness of new vocabulary learning?	
Dose the activity involve negotiation?	
Retrieval	
Dose the activity involve retrieval of the word?	
Is it productive retrieval?	
Is it recall?	
Are there multiple retrievals of each word?	
Is there spacing between retrievals?	
Generation	
Dose the activity involve generative use?	
Is it productive?	
Is there a marked change that involves the use of other words?	
Retention	
Dose the activity ensure successful linking of form and meaning?	
Dose the activity involve instantiation?	
Dose the activity involve imaging?	
Dose the activity avoid interference?	
Maximum score: 18	

The main idea behind the two frameworks (i.e., ILH and TFA) is that the design of a vocabulary learning task establishes the quality of the learning. This is what has been attributed

to the Craik and Lockhart's (1972) levels of processing theory which has been regarded as being fundamental for L2 vocabulary learning (Hulstijn & Laufer, 2001; Laufer, 2005, 2006; Laufer & Hulstijn, 2001; Pulido, 2009; Schmidt, 2001).

The ILH and the TFA frameworks are different in the way they evaluate depth of processing and the values they give to each component. Therefore, it is likely that these frameworks have discrepancies in predicting the effectiveness of tasks or activities designed for L2 vocabulary learning. To further delve into this issue, the present study aimed to investigate the predictive power of these two frameworks (i.e., ILH and TFA) across different L2 vocabulary learning tasks. In other words, the main objective of this study was to reveal that which of these frameworks could provide a greater explanatory power in predicting the effectiveness of different vocabulary learning tasks.

Literature Review

The differences between the ILH and the TFA can lead to different weights given to their components, resulting in variations in prediction as to what vocabulary tasks or activities are more effective in L2 learning (Nation & Webb, 2011). The effectiveness of the ILH and its predictive power has examined in a number of studies (e.g., Hu & Nassaji, 2016; Hulstijn & Laufer, 2001; Keating, 2008). However, there is dearth of research on the predictive power of TFA.

The first empirical research on the notion of the ILH was conducted by Hulstijn and Laufer (2001). The study aimed to explore the predictive power of IL on learning 10 English words by Israeli and Dutch young adult EFL learners. Task 1 was a reading comprehension with marginal glosses with IL of 1; task 2 was reading comprehension plus fill in with IL of 2; and task 3 was writing a composition and incorporating the target words with IL of 3. After doing the tasks, the participants took two vocabulary post-tests. The results showed that task 3 was the most effective vocabulary task which confirmed the ILH. However, tasks 1 and 2 did not show any significant differences despite different IL indices. In another study, Keating (2008) designed three tasks with different involvement load indices to examine the Laufer and Hulstijn's (2001) ILH among low-proficiency Spanish learners. The first task was reading comprehension with marginal glosses (IL index of 1) because understanding the passage and answering the comprehension questions needed the knowledge of the target words, but search and evaluation were absent (1+0+0=1). The second task was a reading comprehension plus fill-in. It induced moderate need, no search and moderate evaluation (1+0+1=2). The third task was writing original sentences using the target words. Need was moderate, search was absent, and evaluation was strong (1+0+2=3). Retention in the second and the third tasks was higher than the first task in both immediate and delayed post-tests. The results revealed that the ILH predictions could be generalized to low-proficiency learners.

In the same line, Yaqubi, Rayati and Allemzade Gorgi (2010) carried out a study on 60 EFL learners. The participants were expected to learn 10 target words through three vocabulary learning tasks with different IL indices. Task 1, an input-oriented task, included reading comprehension plus using dictionary with the IL index of 3; task 2, an input-oriented task, was fill-in-the-blank plus glosses with the IL index of 2; task 3, an output-oriented task, was a reading comprehension and writing a composition plus glosses with IL index of 3. However, the findings did not support the Laufer and Hulstijn's (2001) ILH. Task 2 was better than Task 1 with a higher IL. Moreover, Task 3 was superior to Task 1. Yaqubi et al. (2010) claimed that output-oriented tasks could result in more desirable outcome regardless of the degree of the IL of a task. In a recent study, Tang and Treffers-Daller (2016) arranged six tasks with different involvement loads

to identify the impacts of the tasks and the extent of IL index on L2 vocabulary learning. Task 1 was reading with glosses in text but no comprehension questions afterward with the IL index of 0; task 2 included reading with glosses in text but irrelevant to the comprehension questions with the IL index of 0; task 3 was reading with glossary in the end followed by reading comprehension questions with the IL index of 2 (+need, +search, –evaluation); task 4 was a reading with glosses in the margin relevant to the comprehension questions with the IL index of 2 (+need, –search, +evaluation); task 5 included a reading with glossary in the end which consisted of several options relevant to comprehension questions with the IL index of 3 (+need, +search, +evaluation); and task 6 was a reading with glosses in the margin and making sentences afterwards, with IL index of 3 (+need, –search, ++evaluation). The results confirmed the predictability of the ILH.

Recently, Hu and Nassaji (2016) assigned 96 adult EFL learners to four groups to learn 14 unknown words. Task 1 was a reading with multiple-choice questions; task 2 was a reading and choosing definition; task 3 was a reading plus fill in the blanks; and task 4 was a reading and rewording the sentences. They were different in ranking according to the comparisons between the ILH and TFA prepared by Nation and Webb (2011). Task 1, 2, and 4 had IL index of 3, and technique feature score of 6 (consistency between ILH & TFA). Task 3; however, had involvement load index of 2 and technique feature score of 7 (inconsistency between ILH & TFA). The results from pre-test and post-tests suggested that of the two frameworks, TFA was considerably more powerful than ILH in predicting lexical gains. It was revealed that tasks with higher TFA score (task 3) resulted in better learning.

Considering paucity of research on exploring the predictions permitted by the ILH and the TFA frameworks, this study intended to examine the explanatory power of the two frameworks in predicting the effectiveness of various vocabulary tasks with different IL indices and TFA scores. To fulfill the objectives of the study, the following research question was addressed:

Q: To what extent do Involvement Load Hypothesis and Technique Feature Analysis predict the contribution of L2 Vocabulary Learning Tasks?

Methodology

Participants

Initially, a total of 92 female EFL learners from four intact classes were selected. All participants had been studying English in Iran Language Institute for seven semesters, and they were at pre-intermediate level. However, to ensure the participants' proficiency level, the researchers administered Oxford Placement Test (OPT, Allan, 2004). Those participants whose scores were between 120 and 134, which was equivalent to the pre-intermediate level, were included in this study. Moreover, to assure that participants had a logical degree of word knowledge required for reading a pre-intermediate text, the 2000 Vocabulary Level Test (Schmitt, Schmitt, & Clapham, 2001) composed of the most frequent 2000 words in English was given to them. The participants who satisfied the threshold level of the test (28 out of 30 points on the test), as proposed by Schmitt et al. (2001), were selected. After administering the OPT and Vocabulary Level Test, the number of participants reduced to 80. The participants' age ranged from 19 to 25 ($M = 21.35$, $SD = 1.47$). Afterwards, the participants were randomly assigned to one of the four vocabulary learning tasks (i.e., reading with glosses, keyword technique, word card, reading and finding the words in text). They were also given detailed instructions on how to do the tasks and were assured of the confidentiality of results.

Instruments and Materials

Vocabulary Learning Tasks

The aim of this study was to evaluate the extent to which four vocabulary learning tasks with various rankings given by the ILH and TFA can lead to better vocabulary learning. Nation and Webb (2011) analyzed 12 tasks using the TFA and ILH frameworks. To fulfill the objectives of the study, the researchers opted out four tasks proposed by Nation and Webb (2011) including reading with glosses, using the keyword technique, word card, and finding the words in the text. Following Hu and Nassaji's (2016) study, the tasks were arranged according to the amount of their IL as high, moderate and low. Moreover, the tasks were arranged based on their technique feature score as high, moderate and low. The tasks that received IL indices of 1 and 2 were categorized as having a low IL; the task with the IL index of 3 was categorized as having a moderate IL; and the tasks with the IL index of 4 and 5 were considered as having a high IL. Correspondingly, those tasks that received a score of 5 and below ranked as low scores of TFA; tasks with the score of 8 and 9 were considered as moderate; and tasks that received 11 and above were regarded as high scores of TFA.

Task 1 was reading with glosses with the ILH index of 1 and TFA score of 5. To this end, two passages appropriate for the pre-intermediate level learners were selected from http://www.ngllife.com/. Each passage consisted of eight target words. According to Finocchiaro and Bonomo (1973), no more than about eight new words should be presented at one time; otherwise, it is not manageable by the students. The readability coefficients of the two passages according to Flesch-Kincaid Grade Level were 7.9 and 7.2, respectively, which showed that the readings were appropriate for pre-intermediate learners. Flesch-Kincaid Grade Level is used extensively in the field of education, the formula, is used to judge the readability level of various books and texts. The meanings of the target words in the text were presented through glosses provided in the margin of the text.

Task 2 was keyword technique with the ILH index of 2 and TFA score of 8. In the key word technique task, to help learners remember the meaning of the new words, the language learners were asked to think of a key word from their first language which sounded like the word or the beginning of the new word. In this way, the learners could form a mental image which associated the meaning of the new word to the meaning of the keyword. As the number of the target words was 16, the researchers divided them into two parts based on the words presented in task 1. Each task was presented in one session; therefore, task 2 was conducted in two sessions.

Task 3 was word cards with the ILH index of 3 and TFA score of 11. Learners were given word cards featuring the L2 word on one side with its part of speech and the L1 translation on the other side of the card. Learners went through the cards until they got the meaning of the target words. The task was conducted in two (each session 8 words).

Task 4 included finding the words in the text with the ILH index of 4 and TFA score of 8. Two texts with the target words were given to the participants. After reading the text, the meanings of the target words were provided. Then, the participants were required to search in the passage to find the target words. The rankings of each task suggested by Nation and Webb (2011) are depicted in Table 2.

Table 2. *The Score of Four tasks based on the TFA and ILH Frameworks (adopted from Nation & Webb, 2011)*

	Task 1	Task 2	Task 3	Task 4
Motivation				
Is there a clear vocabulary learning goal?	0	1	1	1
Dose the activity motivate learning?	0	1	1	1

Do the learners select the words?	0	1	1	0
Noticing				
Dose the activity focus attention on the target words?	1	1	1	1
Dose the activity raise awareness of new vocabulary learning?	1	1	1	1
Dose the activity involve negotiation?	0	0	0	0
Retrieval				
Dose the activity involve retrieval of the word?	0	0	1	1
Is it productive retrieval?	0	0	0	1
Is it recall?	0	0	1	0
Are there multiple retrievals of each word?	0	0	1	0
Is there spacing between retrievals?	0	0	1	0
Generation				
Dose the activity involve generative use?	1	0	0	1
Is it productive?	0	0	0	0
Is there a marked change that involves the use of other words?	0	0	0	0
Retention				
Dose the activity ensure successful linking of form and meaning?	1	1	1	0
Dose the activity involve instantiation?	0	0	0	0
Dose the activity involve imaging?	0	1	0	0
Dose the activity avoid interference?	1	1	1	1
Total TFA Score	5	8	11	8
Involvement Load Index (need, search, evaluation)	1+0+0= 1	2+0+0= 2	2+1+0= 3	1+2+1= 4

Target Words and Tests
Initially, 25 words were selected from two reading texts at pre-intermediate level obtained from www.ngllife.com. The 25 words were given to 20 EFL learners who were at the same level of language proficiency as the participants of the main study, and they were asked to translate all the words into Persian. There were 16 words unknown to them. Therefore, these 16 words were selected as the target words. The target words were *endangered, brutal, detour, sprint, leap, slam, gouge, endure, shriek, skid, exploit, posture, insult, profitable, smooth,* and *distract* (8 verbs, 4 adjectives and 4 nouns).

Participants' knowledge of the target words was tested two weeks before the treatment, using a list of the target words which consisted of 16 target wards (8 verbs, 4 adjectives and 4 nouns) for which the participants had to provide Persian equivalents. After completing the tasks, the participants' knowledge of the target words were again tested by the translation task. In the post-test, the order of the target words was changed in order to be different from that of the pre-test. The correct Persian equivalents received the score of 1 and the incorrect ones received 0. The highest possible score was 16 and the lowest one was 0. The inter-rater reliability

coefficients of the pre-test and the post-test measured through Pearson correlation were 1 and 0.97, respectively.

Procedures

Initially, a total of 92 female EFL learners from four intact classes were selected. The researchers administered OPT (Allan, 2004) to ensure the participants' proficiency level. Moreover, to assure that participants had a logical degree of word knowledge required for reading a pre-intermediate text, the 2000 Vocabulary Levels Test (Schmitt et al., 2001) was administered to them. After administering the OPT and Vocabulary Levels Test, the number of participants reduced to 80. Each class was assigned to one of the four vocabulary learning tasks (i.e., reading with glosses, keyword technique, word cards, reading and finding the words in text). Participants' knowledge of the target words was tested two weeks before the treatment, using a list of the target words for which the participants had to provide Persian equivalents.

In reading with glosses group, two passages appropriate for the pre-intermediate level learners were selected. Each passage consisted of eight target words. The meanings of the target words in the passages were presented through glosses provided in the margin of the text.

In word cards technique group, the participants were asked to think of a key word from their first language which sounded like the word or the beginning of the new word. As the number of the target words was 16, the researchers divided them into two parts based on the words presented in reading with glosses group.

In keyword technique group, the participants were given word cards featuring the English word. The participants went through the cards until they got the meaning of the target words.

In reading and finding the words in text task, two texts with the target words were given to the participants. After reading the text, the meanings of the target words were provided. Then, the participants were required to search in the passage to find the target words.

After completing the tasks, the participants' knowledge of the target words were again tested by the translation task. In the post-test, the order of the target words was changed in order to be different from that of the pre-test.

Data Analysis

After collecting the data, a one-way analysis of variance (ANOVA) on pre-test scores was conducted and no significant differences between the four groups were found. Afterwards, a one-way ANOVA was performed to analyze the learners' performance on the post-test. LSD post hoc tests were also carried out to explore the differences among the four tasks. Then, vocabulary gains and weighted scores for the ILH and TFA frameworks were calculated. To measure the participants' weighted scores, their scores in each task were first turned into percentile scores using the following formula: n/16 × 100 = p% (n = score of the task; 16 = the number of target words). Then, the percentile scores were converted into weighted scores based on the number of the components of the ILH and TFA models. Finally, a hierarchical multiple-regression was carried out to determine which of the frameworks (ILH or TFA) could predict higher vocabulary gains from the pre-test to the post-test.

Results

First, the descriptive statistics including mean and standard deviation of the four groups was measured (see Table 3).

Table 3. *Descriptive Statistics of the Pre-test Scores*

Tasks	N	Mean	Std. Deviation

Task 1	20	0.55	0.887
Task 2	20	0.45	0.686
Task 3	20	0.20	0.410
Task 4	20	0.35	0.671
Total	80	0.39	0.684

As displayed in Table 3, the mean score of the groups was very low ($M = 0.39$), indicating that almost all participants were not familiar with the target words. Then, an ANOVA was carried out on the participants' pre-test scores (see Table 4). Before doing ANOVA, the homogeneity of the participants was controlled applying Levene's test and the assumptions were met.

Table 4. *ANOVA on Participants' Pre-test Scores*

	Sum of Squares	df	Mean Square	F	Sig.
Between	1.338	3	.446	.950	.421
Groups	35.650	76	.469		
Within Group	36.988	79			
Total					

p < .05

As Table 4 shows, there was not a significant difference among the groups before the treatment [F (3, 76) = .950, *p* = .421]. It implies that the participants' knowledge of the target words prior to the treatment did not significantly differ.

To answer the first research question, first, four tasks were arranged according to the amount of the involvement as proposed by the degree of task-induced involvement (high, moderate and low) in the ILH and technique feature score proposed by the TFA framework (high, moderate and low) (see Table 5).

Table 5. *Task Classification based on the Indices Provided by the ILH and the TFA*

Tasks	Degree of task-induced involvement load	Degree of TFA
Task 1	Low(1)	Low (5)
Task 2	Low (2)	Moderate (8)
Task 3	Moderate (3)	High (11)
Task 4	High (4)	Moderate (8)

As Nation and Webb (2011) proposed the minimum and maximum score for ILH could range from 0 to 6 based on the three criteria, each of which may be absent, present in a moderate form, or present in a strong form. For TFA, the given score could range from 0 to18 based on the scores given to 5 components and 18 criteria. Table 5 illustrates that task 4 which received a score of 4 by the ILH was ranked as having a high involvement. Task 3 that received a score of 3 by the ILH was regarded as having a moderate involvement. Tasks with 1 and 2 IL index were classified as having a low involvement. Correspondingly, task 3 receiving a score of 11 was considered as a high TFA degree. Tasks 2 and 4 that received the score of 8 were classified as moderate TFA degree, and task 1 with the TFA score of 5 was considered as a low TFA degree. Accordingly, reading with glosses (task 1) was considered as a task with low involvement load index and low technique feature score. Word card task (task 2) was classified as a moderate-involvement load index but high technique feature score. Using keyword technique task (task 3)

was classified as a low-involvement load index and moderate technique feature score. Finally, finding the words in the text (task 4) was considered as high-involvement load index and moderate technique feature score. Table 6 presents descriptive statistics including the mean and standard deviation of the four groups' post-test scores.

Table 6. *Descriptive Statistics of the Post-test Scores*

Tasks	N	Mean	Std. deviation
Task 1	20	6.50	2.856
Task 2	20	12.90	1.804
Task 3	20	11.25	2.235
Task 4	20	8.45	2.235
Total	80	9.78	3.424

Table 6 depicts that participants doing task 2 (keyword technique) had the best mean score in the post-test with a mean of 12.90 and task 1 (reading with glosses) with the mean score of 6.50 had the lowest mean score in the post-test. Before doing ANOVA on the post-test scores, the normal distribution of the post-test scores was checked through Shapiro-Wilk test of normality and it was found that the participants' scores across four groups were normally distributed. Table 7 shows the results of ANOVA on the post-test scores.

Table 7. *ANOVA on Participants' Post-test Scores*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	488.450	3	162.817	28.284	0.000
Within Groups	437.500	76	5.757		
total	925.950	79			

$p < .05$

As Table 7 displays, a statistically significant difference [F (3, 76) = 28.284; p = 0.000] among the performances of the groups on the post-test was observed since the p value was lower than 0.05. To further explore the differences among the tasks, LSD post hoc tests were carried out (see Table 8).

Table 8. *Post hoc Multiple Comparisons across the Four Tasks*

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
Task 1	Task 2	-6.400[*]	.759	.000
	Task 3	-4.750[*]	.759	.000
	Task 4	-1.950[*]	.759	.012
Task 2	Task 1	6.400[*]	.759	.000
	Task 3	1.650[*]	.759	.033
	Task 4	4.450[*]	.759	.000
Task 3	Task 1	4.750[*]	.759	.000
	Task 2	-1.650[*]	.759	.033
	Task 4	2.800[*]	.759	.000
Task 4	Task 1	1.950[*]	.759	.012
	Task 2	-4.450[*]	.759	.000

	Task 3	-2.800[*]	.759	.000

$p < .05$

The results of post hoc analyses indicate that the mean score of task 2 was significantly higher than the mean score of other tasks ($p < .05$) and the mean score of task 1 was significantly lower than the mean score of other tasks ($p < .05$). Following the post hoc analyses, the tasks were compared against the assumptions of the ILH and TFA frameworks (see Table 9).

Table 9. *Cross-task Comparisons Checked against the Assumptions of the ILH and TFA*

Cross-task comparisons	Mean differences (sig.)	Assumptions of the ILH	Assumptions of the TFA
Task 2 > Task 1	6.400[*]	√	√
Task 2 > Task 3	1.650[*]	×	×
Task 2 > Task 4	4.450[*]	×	×
Task 3 > Task 1	4.750[*]	√	√
Task 3 > Task 4	2.800[*]	×	√
Task 4 > Task 1	1.950[*]	√	√

As Table 9 illustrates, from among comparisons with significant differences, two of them were inconsistent with the assumptions of both the ILH and the TFA, that is, the participants doing task 2 (keyword technique) outperformed those doing task 3 (word card) and the participants doing task 2 (key word technique) did better than those doing task 4 (finding the words in text). However, contrary to the assumptions of the ILH and in line with the assumptions of the TFA, task 3 (word card) did better than those task 4 (finding the words in the text) in the post-test. Moreover, three comparisons were consistent with both the ILH and TFA, i.e., task 2 > task 1, task 3 > task 1, and task 4 > task 1. Generally, four comparisons were consistent with the TFA assumptions, but three comparisons were consistent with the ILH assumptions.

Vocabulary Gains and Weighted Scores for the ILH and TFA

The participants' performance on the pre-test and the post-test was considered to determine vocabulary learning gains. The analysis was also calculated to examine which of the two frameworks (i.e., ILH and TFA) accounted for more variance in the participants' vocabulary gains. To this end, weighted scores for the ILH and TFA were computed according to the participants' test scores for each task. Weighted scores were determined applying the weight given to the components of ILH (i.e., need, search and evaluation). Based on the number of the components of the two models, at first learners' test scores for each task were turned into percentile scores and the obtained percentile scores were converted into weighted scores. The following formula was followed to calculate the percentile score for each task: $n/16 \times 100 = p\%$ (n = scores of the task; 16 = the number of target words). Table 10 shows the percentage of distribution of the three components in the ILH.

Table 10. *Percentage of Distribution of the Three Components in the ILH*

Tasks	Need	Search	Evaluation
Task 1	1(100٪)	0	0
Task 2	2(100٪)	0	0
Task 3	2(67٪)	1(33٪)	0
Task 4	1(25٪)	2(50٪)	1(25٪)

Note:
Task 1=100/1 = 100	Task 1: 0-100=1
Task 2= 100/2 = 50	Task 2: 0-50=1; 51-100=2
Task 3=100/3 = 33.333	Task 3: 0-33=1; 34-67=2; 67-100=3
Task 4=100/4 = 25	Task4: 0-25=1; 26-50=2; 51-75= 3; 76-100=4

Based on various components presented by the ILH framework, the percentile score was then distributed into percentile ranks. That is to say, as for ILH, three components suggested by ILH for each task, the scores of each task were turned into percentile ranks. As when a participant doing the forth task acquired a gain score of 7. Therefore, his /her percentile score was 7/16×100=43.75%. On the basis of the three components of ILH, a score of 43.75% falls between 26% and 50% percentile rank which is equivalent to an involvement index of 2 (refer to Note for Table 10). Weighted scores were determined applying the weight given to the components of the TFA (i.e., motivation, noticing, retrieval, generation and retention*)* (see Table 11)*.*

Table 11. *Percentage of Distribution of the Five Components in the TFA*

Tasks	Motivation	Noticing	Retrieval	Generation	Retention
Task 1	0	2(40%)	0	1(20%)	2(40%)
Task 2	3(37.7%)	2(25.1%)	0	0	3(37.7%)
Task 3	3(27%)	2(18%)	4(36%)	0	2(18%)
Task 4	2(25%)	2(25%)	2(25%)	1(12.5%)	1(12.5%)

*Note:*Task 1 =100/5=20
Task 1: 0-20=1; 21-40=2; 41-60=3; 61-80=4; 81-100=5
Task 2 = 100/8=12.5
Task 2: 0-12.5=1; 12.6-25.1=2; 25.2-37.7=3; 37.8-50.3=4; 50.4-62.9=5; 63-75.575.6-88.1=7; 88.2-100=8
Task 3 = 100/11=9
Task 3: 0-9=1; 10-18=2; 19-27=3; 28-36=4; 37-45=5; 46-54=6; 55-63=7; 64-72=8; 73-81=9; 82-90=10; 91-100 =11
Task 4 = 100/8=12.5
Task 4: 0-12.5=1; 12.6-25.1=2; 25.2-37.7=3; 37.8-50.3=4; 50.4-62.9=5; 63-75.5=6; 75.6-88.1=7; 88.2-100=8

For the TFA, the scores of each task were turned to percentile rank according to its five components. If a participant in task 4 received a gain score of 7 then his percentile rank was 7/16×100=43.75. According to the distribution of five components of TFA, 43.75 was placed between 37.8% and 50.3% percentile rank which represents the TFA index of 4 (refer to Note for Table 11).

Results of Hierarchical Multiple Regression Analyses
To distinguish which of the two frameworks' predictions could lead to more amount of gains from pre-test to post-test, a hierarchical multiple regression was carried out (see Table 12). To this end, gain scores from the pre-test to post-test were taken as a dependent variable, and the tasks with either ILH or TFA index scores as predicator variables. In order to compare the contribution of each of the predicator variables over and above the contribution of the other, the two predicator variables were entered into the regression model in different arranges.

Table 12. *Multiple Regression Analysis of Variables Predicting Word Gains*

	R	R^2	ΔR^2	ΔF	df	Sig. F change
Model 1						
1. ILH	.619	.383	.383	48.480	(1, 78)	.000
2. TFA	.891	.793	.410	152.890	(1, 77)	.000
Model 2						
1. TFA	.887	.786	.786	286.546	(1, 78)	.000
2. ILH	.891	.793	.007	2.761	(1, 77)	.101

Note: ΔR^2 = change in R^2; ΔF = change in F.

First, the ILH was entered into the equation, and it justified 38% of the variance in the quantity of gains from the pre-test to the post-test, which was significant. Then the TFA was entered into the equation and it explained 41% additional of variance that was statistically significant. On the contrary, for the second step, first the TFA was entered into the equation. In the circumstances which the TFA was entered at the beginning, it clarified 78% of variance in the quantity of gains which was statistically significant. By the time the ILH was entered on the next step, the ILH formed only 0.7 % additional of the variance, which was not statistically significant. The outcomes of the regression analyses revealed that of the two frameworks (ILH and TFA) as predicator variables, the TFA was a stronger predicator for vocabulary gains than the ILH which means that the TFA had a stronger explanatory power in accounting for vocabulary gains.

Discussion

The results revealed that different tasks with different ILH indices and TFA scores had different effects on the participants' vocabulary learning which is in line with a number of studies (Laufer & Hulstijn, 2001; Keating, 2008). There can be some reasons behind obtaining these results. For instance, task 1 that was reading with glosses with the ILH index of 1 had the weakest outcome. It could be because of no search and evaluation in task 1 and the need score of 1 which is low. Therefore, it was expected that the outcome of the post-test in this kind of task be in line with some previous studies (e.g., Hulstijn & Laufer, 2001; Keating, 2008; Tang & Treffers-Daller 2016) which claimed that word learning in an L2 is contingent upon a task's IL (i.e., the amount of need, search, and evaluation it imposes). Moreover, regarding the TFA score, there was no clear vocabulary learning goal for this task; there was no negotiation in this task and it did not involve retrieval of the words. Moreover, it did not motivate learning. According to Nation and Webb (2011), this task could not be productive because it did not involve instantiation and imaging. In addition, the results indicated that the tasks with TFA scores of 8 and 11 (tasks 2, 3, 4) contributed to better and more satisfactory post-test scores than the task with TFA score of 5 (task 1). The results could be because of the motivation weighed score of 3 or 2 for these three tasks in contrast to task 1 with the motivation weighed score of 0. In TFA, *motivation* has a key role like Laufer and Hulstijn's (2001) *need* factor which would get a score of 1 when the attention to the word is as a result of the learners' interest (word card, keyword technique and reading and finding the words in context). Task 3 with the TFA score of 11, which was the highest score, led to a higher post-test performance than tasks 1 and 4. However, task 2 with the TFA score of 8 outperformed task 3. Comparing retention score of the two tasks, task 2

with the score of 3 performed better than task 3 with the retention score of 2. Moreover, the other reason could be the imageability of the keyword technique which led to the better learning of the target words. Furthermore, keyword technique makes use of the very powerful condition of visualization and linking it to remembering the form of the word which can facilitate the process of L2 vocabulary learning (Nation & Webb, 2011).

The analysis of hierarchal multiple regressions also showed that TFA made stronger contribution than the other variable (ILH). TFA alone accounted for 78% of variance in quantity of gains and it was statistically significant. But ILH formed only 7% of variance which was not statistically significant. In other words, ILH did not explain a significant proportion of vocabulary gains. This suggests that TFA is a stronger predicator for vocabulary gains than the ILH. The findings are in line with Hu and Nassaji's (2016) study that showed the higher predictability of the TFA framework.

TFA came into the view to propose more responsive factors which measured vocabulary learning. In this study, tasks 2 and 3 generated better post-test performance. By inspecting the distribution of the five factors within the TFA between the two tasks, it was detected that both of them covered motivation, noticing and retention but absence in generation. It could be reasonably argued that those three factors were conducive to the participants' initial vocabulary learning. Among tasks, task 3 (word card) in post-test performed better than task 1 and 4, it could be because the learners went through the cards at increasingly spaced intervals until the meaning of the words were known, which was receptive learning. By turning the cards over and looking at the translations and trying to recall the 16 word forms, productive learning could occur which generated better performance in post-test of task 3 than tasks 1 and 4. This could be explained that productive components of the tasks may have helped the aspect of the link between the meaning and the form of the target words. Therefore, it could have raised the learning (Nation & Webb, 2011). In this regard, the findings of this study were in line with Hu and Nassaji's (2016) study which claimed that productive components would not only help learners better notice their gap of knowledge (Swain, 2005), but it could be helpful for vocabulary learning (Keating, 2008; Laufer, 2005, 2006). Generally, it can be argued that compared to the ILH, the TFA framework with 5 components and 18 criteria can better predict vocabulary learning gains. In addition, the study did not confirm the assumption for ILH that tasks with sequential ILH index (1, 2, 3 and 4) can lead to the sequential vocabulary gains. This might be because of different weights of the three components presented in the ILH which needs more research in the future.

Conclusion

To conclude, the present research has empirically suggested that of the two frameworks, namely ILH and TFA, the latter can be a more powerful predicator of vocabulary learning. Moreover, motivation was found to play an important role in vocabulary learning. The results also showed that successful retrieval also contributes to learning and increasing the number of retrievals may have a positive effect on learning gains (Nation & Webb, 2011).

Pedagogically speaking, vocabulary activities may assist teachers to enhance learners' vocabulary gains. Teachers can arrange activities to help students develop their vocabulary learning through tasks such as keyword technique, word card, reading and finding the words in text with high or moderate TFA score. Language teachers should provide opportunities for learners to learn more words through more effective tasks. The results also suggested that the tasks consisting of multiple retrievals of words contributed to the participants' vocabulary learning, and the amount of the number of retrievals raised the learning vocabulary gains.

Therefore, EFL teachers are advised to use the tasks with the retrievals of words in their teaching programs to get better results (Keating, 2008; Laufer, 2005, 2006).

The results of the current study are not devoid of limitations. First, it was not possible for the researchers to randomize the learners. Therefore, the intact classes were selected which can be considered as a hurdle to generalize the result of the study. The indices applied for the tasks were based on the Nation and Webb's (2011) suggestions. In this study, there were four tasks with different involvement indices varying from 1 to 4 which were sequential. Therefore, future research could apply more tasks with similar and different ILs, which perhaps have different results in vocabulary gains. Next, more qualitative observation and accurate analysis of vocabulary learning activities are needed to explore whether their use matches their goals to suggest some helpful guidelines to adapt the techniques. Finally, in this study, only pre-intermediate EFL learners were examined. Hence, future research could be administered on some other language proficiency levels.

References

Allan, D. (2004). *Oxford placement test.* Oxford: Oxford University Press.

Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System, 53*(1), 84-95. doi: 10.1016/j.system.2015.07.006

Craik, F. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. doi: 10.1016/S0022-5371(72)80001-X

Finocchiaro, W. S., & Bonomo, M. (1973). *The foreign language learner: A guide to teachers.* New York: Regents Publishing Company.

Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System, 56*(1), 28-39. doi: 10.1016/j.system.2015.11.001.

Hulstijn, J. H, & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558. doi: 10.1111/0023-8333.00164.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*(3), 365-386. doi: 10.1177/1362168808089922

Knight, S. M. (1994). Dictionary use while reading: The effect on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal, 78*(3), 385-299. doi: 10.1111/j.1540-4781.1994.tb02043.x

Laufer, B. (2005). Focus on form in second- language vocabulary learning. In S. H. Foster-Cohen, M. Garcia-Mayo, & J. Cenoz (Eds.) *Eurosla yearbook* (pp. 223-250*).* Amsterdam: John Benjamin's.

Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review, 63,* 149-166*.* doi: 10.3138/cmlr.63.1.149

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26. doi: 10.1093/applin/22.1.1

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

Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary.* Boston: Heinle.

Pulido, D. (2009). How involved are American L2 learners of Spanish in lexical input processing tasks during reading? *Studies in Second Language Acquisition, 31*(1), 31-58. doi: 10.1017/S0272263109090020

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32); Cambridge, UK: Cambridge University Press.

Schmitt, N. Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Level Test, *Language Testing, 18*(1), 55-88. doi: 10.1177/026553220101800103

Swain, M. (2005). The output hypothesis: theory and research. In E. Heinkel (Ed.), *Handbook of research in second language teaching and learning* (pp.471-483). Mahwah, NJ: Lawrence Erlbaum Associates.

Tang, C., & Treffers-Daller, J. (2016). Assessing incidental vocabulary learning by Chinese EFL learners: A test of the involvement load hypothesis. In G. Yu, & Y. Yin (Eds.) A*ssessing Chinese learners of English* (pp. 121-149). UK: Palgrave Macmillan.

Yaqubi, B., Rayati, R. A., & Allemzade Gorgi, N. (2010). The involvement load hypothesis and vocabulary learning: the effect of task types and involvement index on L2 vocabulary acquisition. *Journal of Teaching Language Skills, 2*(1), 145-163. doi: 10.22099/jtls.2012.404