# Task Effectiveness Predictors: Technique Feature Analysis versus Involvement Load Hypothesis

[1] Hooshang Khoshsima*

[2] Zahra Eskandari

## Abstract

How deeply a word is processed has long been considered as a crucial factor in the realm of vocabulary acquisition. In literature, two frameworks have been proposed to operationalize the depth of processing, namely the Involvement Load Hypothesis (ILH) and the Technique Feature Analysis (TFA). However, they differ in the way they have operationalized it specially in terms of their attentional components. The present study made attempts to compare the predictability of these two frameworks for foreign language vocabulary learning task effectiveness. Seventy-six adult EFL learners in Chabahar Maritime University were randomly given one of the four vocabulary learning tasks which were ranked differently by the two frameworks and were required to learn the meaning of 10 target words. The results of the study revealed that TFA had a better explanatory power in predicting vocabulary learning gains than the ILH. The results have implications for language teachers, material developers and syllabus designers.

*Keywords:* Technique Feature Analysis, Involvement Load Hypothesis, Vocabulary Acquisition

## 1. Introduction

Vocabulary learning is at the heart of language acquisition, no matter if the language is first, second or foreign language (Decarrico, 2001; Mobarge, 1997). There is now a consensus among vocabulary specialists that lexical competence highly correlates with communicative competence, the ability to communicate successfully and appropriately (Coady & Huckin, 1997). Vocabulary knowledge seems to be the most clearly identifiable subcomponent of the ability to read (Cain & Oakhill, 2011; Laufer, 1997). Laufer (1997) has come to the conclusion that "the threshold for reading comprehension is, to a large extent lexical" (p. 21). Having a vast store of vocabulary knowledge has also been found to have a significant effect on writing quality (Lee, 2003). It is also one of the major indicators of language learners' proficiency level (Nation, 2001; Zou 2016b). Nonetheless, how vocabulary is learned or what processes are involved has been the focus of much theoretical discussion (Laufer & Hulstijn, 2001; Nation & Webb, 2011).

One closely related debate is that in order to learn vocabulary effectively, learners must deeply process different aspects of words (Hu & Nassaji, 2012; Hulstijn & Laufer, 2001; Laufer, 2005; Nassaji, 2003, 2004; Hu & Nassaji, 2012; Schmidt, 2001). This is referred to as elaborate processing or depth of processing framework and was first proposed by Craik and Lockhart (1972) and has been emphasized to be essential for L2 vocabulary learning (Ellis, 1994; Hulstijn & Laufer, 2001; Laufer, 2005, 2006; Laufer & Hulstijn, 2001; Pulido 2009; Schmidt 2001). It holds that "the memory trace can be understood as a by-product of perceptual analysis and that trace persistence is a positive function of the depth to which the stimulus has been analyzed" (p. 671). Based on this framework, the deeper the processing of a stimulus is, the traces in memory will be more elaborate, longer lasting and stronger. The hypothesis suggests that the

[1] Associate Professor, khoshsima2002@yahoo.com; Department of English Language, Chabahar Maritime University, Chabahar, Iran

[2] PhD Candidate, eskandari62@gmail.com; Department of English Language, Chabahar Maritime University, Chabahar, Iran

retention of information is determined by the depth in which it is processed rather than the length of time it is held in the primary memory. They have also posited several levels of processing depth. For example, processing the semantic features of a lexical item (e.g., meaning) is supposed to occur at a deeper level than its structural features (e.g., orthography). In other words, tasks which require the learners to process the meaning of words lead to better word retention.

One of the main problems associated with the depth of processing hypothesis was the lack of operationalizable definition, based on which tasks could be graded and evaluated in terms of their processing depth and effectiveness. To tackle the problem, two frameworks have been proposed in literature as effective ways to operationalize the levels of processing theory: *Involvement Load Hypothesis* (Laufer & Hulstijn 2001) and *Technique Feature Analysis* (Nation & Webb, 2011).

The present study was designed to examine and compare the predictions yielded by the two frameworks. The aim was to find out which of the two frameworks provided a greater explanatory power in predicting the effectiveness of different vocabulary learning tasks.

## 2. Literature review

### 2.1. The Involvement Load Hypothesis

In an attempt to operationalize the depth of processing model which was put forward by Craik and Lockhart (1972), and also to account for the variation in effectiveness among different vocabulary learning tasks, Laufer and Hulstijn (2001) proposed a cognitive-motivational framework called the *involvement load hypothesis* according to which tasks that induce higher involvement load are conducive to the type of processing that is considered vital to vocabulary retention. The construct of the involvement has three components: *need*, *search*, and *evaluation*. Each of these components can be either present or absent while processing a word during a task; and in case they are present, each can be either moderate or strong. *Need*, which is the motivational component is deemed *strong* or *moderate* depending on whether the learner is intrinsically or extrinsically motivated respectively. For example, it is moderate when the teacher wants the learner to find the meaning of a word. However, it is strong when it is self-imposed by the learner (e.g. looking up a word in a dictionary while reading a text). Search could be either moderate or strong depending on whether it is receptive retrieval or productive retrieval (Nation & Webb, 2011). If the learner searches for the L2 meaning, it is moderate and if the learner searches for the L2 form it is strong. Apropos of evaluation, it is moderate when the learner compares the particular meaning of a word with other meanings; yet, it is strong when the learner has to see if a word meaning can fit a specific linguistic context or not. In a task, according to Laufer and Hulstijn (2001), each of these components could be 0 ,1 or 2 depending on whether they are absent, moderate or strong respectively. The combination of these factors with their degrees of prominence constitutes the *task induced involvement load*. The higher the involvement load, the more effective the vocabulary task is.

To further illustrate the construct of *involvement index,* two tasks with varying involvement loads are presented alongside as an example (Hulstijn & Laufer, 2001). In the first task, the learner is asked to write original sentences with some new words whose meanings are provided by the teacher. In this case, the need is moderate (imposed by the teacher), there is no search (meanings are provided), and strong evaluation is required in that the learner has to use the new words in learner-generated contexts. In view of the involvement indexes, the task can be described as having an involvement index of 3 (1 + 0 + 2). In the second task, the learner is required to read a text (with glosses of the new words) and to answer comprehension questions. The task thus induces a moderate need, yet neither search nor evaluation. The involvement index of this task is accordingly 1 (1 + 0 + 0). The first task is postulated, according to the construct of task-induced involvement, to induce a greater involvement load than the second task. With the preliminaries of ways of measuring involvement load, the basic contention of the involvement load hypothesis is that "retention of unfamiliar words is, generally, conditional upon the degree of involvement

in processing these words" (Hulstijn & Laufer, 2001, p. 545). In other words, the researchers argue that the greater the involvement load, the better the retention.

The hypothesis has some basic assumptions. Hulstijn and Laufer (2001) consider time-on-task as an inherent feature of the task not amenable to manipulation arguing that differences in completion time reflect inherent differences in task demands (Keating, 2008). The hypothesis also postulates that none of the involvement factors gets priority over another and no particular task type - be it input task or output task– is considered superior or more effective (Keating, 2008). The only determinant factor, based on the hypothesis, is the degree of involvement that a task induces.

Since the introduction of the involvement load hypothesis, many studies have tried to investigate its efficacy from different aspects. Some have tried to examine the effect of tasks with different involvement loads on learning new words and have obtained evidence for it (Bao, 2015; Ghabanchi et al. 2012; Hulstijn & Laufer, 2001; Karalik & Merc, 2016; Keating, 2008; Kim, 2008; Lee, 2003; Reynolds, 2014; Soleimani & Rahmanian, 2014; Soleimani & Rostami Abu Saeedi, 2015; Rott 2005; Sarbazi, 2014; Silva & Otwinowska, 2017; Soleimani et al., 2015; Tahmasbi & Farvardin, 2017; Zou, 2016a). Some others have found evidence against it (Folse 2006; Jahangiri & Abilipour, 2014; Maftoon & Haratmeh, 2012; Martínez-Fernández, 2008; Soleimani & Rahmanian, 2014; Yaqubi et al., 2010).

Hulstijn and Laufer (2001) conducted two parallel experiments in two countries to test their hypothesis. In their study, the participants were provided with three tasks with different involvement loads. The results indicated that the writing condition yielded significantly higher retention than the fill-in and gloss conditions in both experiments, while the fill-in group showed significantly higher retention than the gloss condition in one experiment but not in the other.

Ghabanchi, et al, (2012), partially replicated the study conducted by Hulstijn and Laufer (2001) and found supporting evidence for the involvement load hypothesis. The results of the study were in line with the results of the study conducted by Hulstijn and Laufer (2001) showing that the higher level of learner involvement during tasks promoted more effective initial learning and better retention of new words.

In the study conducted by Folse (2006), the participants practiced the new words under three different conditions: one fill-in-the-blank exercise, three fill-in-the-blank exercise, and one original sentence writing. The results revealed that the three tasks were significantly different from each other, with the words practiced in three fill-in-the-blank exercise retained much better than those practiced under either of the other two tasks. According to the findings of the study, the number of word retrieval and NOT the depth of word processing was the determinant factor in task efficacy. The results were in contrast to the claims made by the involvement load hypothesis. Similarly, Keating (2008) came to the conclusion that when time-on-task was taken into account, "the benefit associated with more involving tasks faded" (Keating, 2008). Likewise, Jahanbini and Abilipour (2014) argued that when time is hold constant, the type of exercise does not make any significant difference in the retention of new words.

To extend the line of the research on the involvement hypothesis, some studies have examined whether tasks with the same involvement load are equally conducive to vocabulary retention (Bao, 2015; Ghabanhi et al., 2012; Kim, 2008; Zou, 2016a). Kim (2008) investigated whether two tasks (i.e., writing a composition and writing sentences) hypothesized to represent the same level of task-induced involvement would result in equivalent initial learning and retention of 20 adult ESL learners at two different levels of proficiency. The study provided some evidence that tasks with the same involvement load were equally beneficial for vocabulary learning. Nonetheless, Kim suggested that more research should be conducted on different degrees of each individual component because they might not contribute to the same weight and that strong evaluation might be the most influential factor in vocabulary acquisition (Kim, 2008, as cited in Hu & Nassaji, 2016). Ghabanchi et al.'s (2012) findings were also in line with Kim's (2008) study;

however, quite the contrary, Zou (2016a) found that composition writing was significantly more effective than sentence writing, despite having the same involvement load.

## 2.2. Technique Feature Analysis

Nation and Webb (2011) have criticized the Involvement Load Hypothesis on the ground that the three components of need, search, and evaluation do not allow the consideration of other related factors that can affect the effectiveness of vocabulary learning activities. Inspired by the ILH, Nation and Webb (2011) developed the Technique Feature Analysis (TFA) as a framework to evaluate task effectiveness in vocabulary acquisition. It was intended to compensate for the inadequacy associated with the ILH by introducing more criteria to operationalize the depth of processing model than those included in the ILH (Hu & Nassaji, 2016). It is in fact a modified version of the vocabulary-learning framework proposed by Nation (2001), suggesting that vocabulary learning entails three components of noticing, retrieval, and generation (Hu & Nassaji, 2016). TFA was developed by adding two more components: motivation and retention. TFA is based on the statement that "the design of the task determines the quality of the learning outcome" (Nation & Webb, 2011, p.4). the purpose of the framework is both to evaluate and design techniques.

So, as mentioned above, TFA includes five components (i.e. motivation, noticing, retrieval, creative use, and retention) and some criteria to assess each component. The questions have been classified based on the psychological conditions contributing to vocabulary learning. The answer to each question is scored as 0 or 1 with the total score indicating the relative value of that activity. The highest score possible is 18. What follows is a brief explanation of each component based on Nation and Webb (2011).

### Motivation

One of the good characteristics of a vocabulary learning activity is having a clear goal and encouraging students to achieve that goal. In other words, they motivate students to do them. Enjoyable activities (e.g. crossword puzzles and games), challenging activities (e.g. cloze activities) and those that raise awareness can all be motivating. This component is also similar to the "need" component of the ILH. However, in TFA, if students are extrinsically motivated, the index of selection would be 0 and if they are intrinsically motivated the index would be 1.

### Noticing

Vocabulary activities should somehow attract learners' attention to the unknown words or to the features of the words that are unknown (e.g. through highlighting or underlining, glossing, etc.). furthermore, an activity should raise the learners' awareness to notice that there is something to learn by, for example, making the learners use words in context, selecting the correct word form from a number of choices and so forth. Besides, an activity would be much more effective if it involves negotiation.

### Retrieval

Nation and Webb (2011) have distinguished several retrievals: *receptive* vs. *productive*, *recognition* vs. *recall*, *multiple retrieval* vs. *single retrieval*, and spaced retrieval vs. massed retrieval. Receptive and productive retrieval is similar to the "search" component in the ILH. Productive retrieval involves trying to find a word form while receptive one involves searching for meaning of a word. "Recall" is different from recognition in that in the latter, the learners are provided with some choices through hearing or seeing from among which the learner can recognize the intended meaning or form while in the former, the learners need to retrieve the word form or meaning from memory. Productive retrieval and recall are considered more difficult than recognition or receptive retrieval. Moreover, multiple and spaced retrievals are deemed to be more useful.

"Meeting word in a new way (receptive generative use) or using a word in a way that the learner has not met before" (Nation & Webb, 2011, p. 9) can strengthen memory and the latter is considered more demanding especially when it involves a marked change. Put it differently, productive generation may take place in different degrees ranging from no generation to low, reasonable and high generation. The difference in degrees is relevant to the amount of creativity and change associated with using a word.

*Retention*

An activity would receive a point if it makes the learner successfully link form and meaning. It would get more credits if it involves seeing a word as it is used in a meaningful situation, imaging, and avoiding interference. Interference occurs when learners have to learn a group of semantically related words. That is, learning semantically unrelated words is easier than learning semantically related ones.

As noted earlier, a number of recent studies have examined the effectiveness of the ILH and have found some evidence for or against its predictive power. However, To the best of the researchers' knowledge, only two studies have so far examined the predictive power of the TFA which were conducted by Hu and Nassaji (2016) and Jafari Gohar, Rahmanian and Soleimani (2018). Hu and Nassaji (2016) made attempts to empirically compare the predictive power of the two frameworks for effective vocabulary learning tasks. In their study, 96 adult EFL learners, who were Taiwanese college-level second-year business majors, were divided into four groups and each group performed one of the vocabulary learning tasks with different ILH and TFA indexes: *reading a text and multiple-choice items on text*, *reading a text and choosing definitions*, *reading plus fill-in-the-blanks*, *and reading a text and sentence rewording*. The results of their study revealed that the TFA had a better explanatory power in predicting task effectiveness than the ILH both in during task performance and in pretest to posttest vocabulary gain.

In the study conducted by Jafari Gohar et al. (2018), 90 high proficiency EFL students were assigned into three vocabulary tasks of *sentence making*, *composition*, and *reading comprehension*. The results of their study were just partially in line with the findings reported by Hu and Nassaji (2016). Jafari Gohar et al. (2018) concluded that the ILH cannot be a good predictor, and the TFA was a good predictor in pretest to posttest score change but not in during task activity.

Considering the somehow contradictory results obtained by the two above-mentioned studies, the present study was conducted to once again investigate and compare the predictions made by these two frameworks, especially to see which of the two models presents greater explanatory power in predicting the efficacy of different vocabulary tasks. The researchers selected the four vocabulary learning tasks used by Hu and Nassaji (2016) which differ in their ranking and the extent to which they promote the different components of the ILH and TFA. First the features of these tasks were examined and their indexes based on the two frameworks were calculated and then their effectiveness in terms of the participants' vocabulary knowledge gains was compared. The following two research questions were examined:

1. To what extent do the four vocabulary tasks used in this study contribute to L2 vocabulary learning?

2. To what extent is the contribution of the four tasks, if any, predicted by the ILH versus the TFA frameworks?

## 3. Method

*3.1. Participants*

A total of 100 adult EFL learners participated in the study. Participants, whose native language was Persian, were students of English as a Foreign language at Chabahar Maritime University in Iran. The sample included both males and females ranging from 19 to 25 years old. They were given an Oxford

Placement Test and those whose scores were one standard deviation above or below the mean were selected to participate in the study: a total of 76. All of the participants were from four intact classes and the data were collected during their regularly scheduled class periods. Within each class, the participants were randomly assigned to one of the four experimental tasks: reading a text and multiple-choice questions on the text (n= 22), reading and fill in the blanks (n= 20), reading a text and sentence rewording (n= 18), and reading and choosing definition (n =16); and those who have been excluded from the study were given a reading comprehension task as placebo.

### 3.2. Target Words

Using the AWL highlighter (Coxhead, 2000), twenty low frequency words were selected from a text decided for the study. Of the twenty words, ten were chosen for the investigation through a pilot study with a similar pool of participants. Moreover, the results of the pretest of the main study showed that the words were unknown to the participants. Of the target words six were nouns, three were verbs and one was adjective.

### 3.3. The Reading Text

The reading passage was adopted from the article in a reading comprehension book (Richeck, 1993). The reading text and the procedures used to modify it were similar to what had been conducted by Ghabanchi et al. (2012). The passage was about the origins of superstitions. The text contained 551 words and it was supposed that the participants would have some general idea of the topic. Except for the target words, the reading was modified so that the vocabulary was kept within a first and second thousand vocabulary list (Nation, 1984). Reducing the number of unknown words in the text frees up the amount of cognitive space required to attend to the massage (Joe 1998, as cited in Ghabanchi, et al., 2012). Thus, as additional resources are made available during text processing, the forging of stronger form-meaning connections is made possible, such that target words may be retrievable at a later time (Craik & Lockhart, 1972; Craik & Tulving, 1975; Laufer & Hulstijn, 2001; Nation, 1990). Another criterion for modifying the text was the number of occurrence for each target word. The passage was revised in such a way that all target words would appear only once.

Some teachers are uncomfortable with simplification largely because they feel that the authenticity of the text is lost. But as Nation (2001) explicitly asserts, authenticity lies in the readers' response to the text and not in text itself. The text was then given to two educated native speakers and two experienced English teachers in Chabahar Maritime University to review. The appropriateness of the revised text difficulty level was also confirmed by the teachers and also through a pilot study. The teachers confirmed that except for the target words, the students knew all the other words in the text and that they would not encounter any of the target words during the semester.

### 3.4. Tasks

As mentioned earlier, the aim of this study was to investigate the extent to which the vocabulary tasks with similar and different rankings between the ILH and TFA can be conducive to vocabulary learning. To this end, four vocabulary tasks which differed both in their rankings and in the extent to which they promote the different components of the ILH and the TFA were developed. Accordingly, the researchers had to select tasks consistent with both frameworks. The tasks for comparison were those suggested in Nation and Webb (2011) and included the following: reading and fill-in-the-blank (Task 1), reading and rewording the sentences (Task 2), reading and choosing definitions (Task 3), and reading and multiple-choice on text (Task 4). Task one had an ILH index of 2 and a TFA score of 7 and the other tasks had an ILH index of 3 and a TFA score of 6. It was not possible to choose tasks with a larger gap of involvement load because the maximum involvement load for the tasks listed in Nation and Webb (2011) is 4.

*Reading and multiple-choice on text*: Learners performing this task were provided with a text and ten multiple-choice comprehension questions based on the reading passage. These questions either incorporated some target words or paraphrased the original sentences in which these target words occurred. Accordingly, the successful completion of the questions entailed understanding of the target lexical items. In the reading passage the ten target words, whose understanding was relevant to the task, were highlighted in bold print. Example:

*Answer the following questions according to the passage*

*Nowadays, most people do not believe superstitions because …………*

> a) *science has developed*
> b) *they have a smattering of science*
> c) *they are in a quandary*
> d) *others might guffaw at them*

*Reading and choosing definitions:* In this task, the target words were highlighted in bold print. Upon finishing reading the text, the participants were required to choose the correct definition of each target word from among four choices (Nation & Webb, 2011, p. 322). Example:

*Choose the correct definition for each word.*

*Slop          a. play          b. pour          c. drive          d. start*

*Reading plus Fill-in-the-blank*: Students in this group were given the same text and the same questions as those in *Reading and multiple-choice on text* group. For this group, however, the ten target words were deleted from the text, leaving ten gaps numbered 1-10. The ten target words, along with one extra word that had not been appeared in the original text, were printed in random order as a list on a separate page with their L1 translation, their L2 explanation and their grammatical function. The task was to read the text, fill in the ten gaps with the missing words from the list of words, and answer the comprehension questions.

*Reading and sentence rewording:* "In this task, the target words were highlighted in the text. Upon finishing reading the text, the learners had to rewrite the sentences drawn from the text containing the target words." (Nation & Webb, 2011, p. 322).

*Reword the sentences without changing the meaning. Use an appropriate form of the words in parentheses if necessary.*

*He poured some salt on the food. (slop)*

A summary of the tasks along with their TFA and ILH indexes are presented in Table 1.

Table 1: Four Tasks Analyzed Using TFA and ILH (Adopted from Hu & Nassaji, 2016)

| Criteria | Reading and multiple-choice on text | Reading and choosing definitions | Reading and fill in the blank | Reading and sentence rewording |
|---|---|---|---|---|
| **Motivation** | | | | |
| Is there a clear vocabulary learning goal? | 0 | 1 | 1 | 1 |
| Does the activity motivate learning? | 1 | 1 | 1 | 0 |
| Do the learners select the words? | 0 | 0 | 0 | 0 |
| **Noticing** | | | | |
| Does the activity focus attention on the target words? | 1 | 1 | 1 | 1 |
| Does the activity raise awareness of new vocabulary learning? | 0 | 1 | 1 | 1 |
| Does the activity involve negotiation? | 0 | 0 | 0 | 0 |
| **Retrieval** | | | | |
| Does the activity involve retrieval of the word? | 1 | 1 | 0 | 0 |
| Is it productive retrieval? | 0 | 0 | 0 | 0 |
| Is it recall? | 1 | 0 | 0 | 0 |
| Are there multiple retrievals of each word? | 0 | 0 | 0 | 0 |
| Is there spacing between retrievals? | 0 | 0 | 0 | 0 |
| **Generation** | | | | |
| Does the activity involve generative use? | 1 | 0 | 1 | 1 |
| Is it productive? | 0 | 0 | 0 | 1 |
| Is there a marked change that involves the use of other words? | 0 | 0 | 0 | 0 |
| **Retention** | | | | |
| Does the activity ensure successful linking of form and meaning? | 0 | 0 | 1 | 1 |
| Does the activity involve instantiation? | 0 | 0 | 0 | 0 |
| Does the activity involve imaging? | 0 | 0 | 0 | 0 |
| Does the activity avoid interference? | 1 | 1 | 1 | 0 |
| **total score** | **6** | **6** | **7** | **6** |
| **Involvement load index (need, search, evaluation)** | 1+1+1=**3** | 1+1+1=**3** | 1+0+1=**2** | 1+0+2=**3** |

*3.5. Pretest and posttest*

A modified version of vocabulary knowledge scale (Paribakht & Wesche, 1997) developed by Folse (2006) was used for both the pretest and posttest. This modified version of the VKS includes three levels of word knowledge that could detect even the partial gains in degrees of knowledge. On this modified scale, one point was awarded if the correct meaning was provided (as evidenced by an acceptable English synonym. English definition, L1 translation or definition). One additional point was

also awarded if the student could make a correct sentence with the word. Thus, each word could receive a score of 0, 1, or 2.

*3.6. The study design and the pre- and posttest measures*

Two weeks prior to the main study, the participants in the four intact classes attended the Oxford Placement Test (OPT). It was conducted during their regular class times. Then those who scored one standard deviation below the mean or one standard deviation above the mean were selected to participate in the study. The researcher then, prepared the tasks. The main text was also revised for its length and complexity. The text was reviewed by two educated native speakers and two English teachers in the university. They all admitted the words were of low frequency. The teachers also acknowledged that the students wouldn't encounter the words during the semester. All tasks along with the pretest and posttest took place in the learners' regular class times on scheduled review days. Although it was assumed that the target words were unfamiliar to the learners, all participants were still given a vocabulary pretest measuring their knowledge of the target words prior to performing the tasks. Randomization of the experimental tasks in this study occurred within groups. The tasks were photocopied and collated into one stack. participants in each of the classes were given one of the four experimental tasks drawn from the top of the collated stack. The researcher visited a total of four intact classes and followed the same administration procedure in each. Other students in each class, who were not of the intended level of language proficiency, were given a reading comprehension task as placebo. Upon the completion of the tasks, the worksheets were collected and the students were unexpectedly given an immediate posttest designed to measure their initial vocabulary learning. The order of the target words in the pretest and posttest were not the same. In addition to measuring the learners' knowledge of the vocabulary items in the pretest and the posttest, their during-task success was also measured. This was done by checking the participants' responses to the target words when they had to perform each of the four tasks. For scoring the during-task performance, similar to Hu and Nassaji (2016), the learners were awarded a score of 1 for a correct response and 0 for an inaccurate answer. For task 4 (i.e., reading a text and sentence rewording), the same scoring system was utilized to evaluate the accuracy of the reworded sentences. They received a score of 1 for a grammatically correct sentence containing the synonym of the original target word, and a score of 0 if the answer was wrong. Two independent raters read and judged their rewritten sentences, and an inter-rater reliability of .98 was also achieved.

**4. Results**

In order to explore the research questions, first the learners' pretests were examined. Except for one participant who knew three of the target words and was excluded from the study, no one knew any of the target words and the participants all scored 0 in their pretest showing that they were all at the same level in terms of their knowledge of the target words before the treatment.

The first research question investigated the extent to which the tasks with similar and different rankings between the ILH and TFA contributed to vocabulary learning. To address the question the tasks were first classified according to the indexes suggested by both the degree of task-induced involvement (high and low IL) in the ILH and the technique feature score in the TFA framework (high and low TFA). Hence, the tasks with a score of 3 by the involvement load were classified as having high involvement and the ones that received a score of 2 were classified as lower involvement. likewise, those tasks that received a score of 7 were classified as high TFA and the ones with a score of 6 were classified as lower TFA. Accordingly, reading and multiple-choice on a text, reading and choosing definitions, as well as reading and sentence rewording tasks were classified as tasks with high-involvement load indexes but lower technique feature scores. However, reading and fill in task was classified as having low involvement load index but a higher technique feature score.

In order to examine the students' differences in task performances across the four tasks, their during-task performances (their correct responses to the target words when performing the task) as well as their vocabulary gain from pretest to posttest were calculated through one-way ANOVA. As Table 2 shows, the mean score of during task performance is the highest in task 1(reading and fill-in-the-blank) with a mean of 7.8000, followed by tasks 2 and 3 with the mean scores of 7.2222 and 7.0000 respectively; and the lowest mean score belongs to task 4 (reading and multiple-choice on text with a mean of 6.2727). A one-way ANOVA was then conducted to examine whether there were any statistically significant differences across the four tasks (Table 3). No statistically significant difference was found among the four tasks. Hence none of the assumptions of the ILH or TFA were met.

Table 2: Descriptive Statistics of the During-task Performance per Condition

| Tasks | N | Mean | SD |
|---|---|---|---|
| Reading and Fill-in-the-blank(Task 1) | 20 | 7.8000 | 1.43637 |
| Reading and sentences rewording (Task 2) | 18 | 7.2222 | 1.86470 |
| Reading and choosing definitions (Task 3) | 16 | 7.0000 | 2.42212 |
| Reading and multiple-choice questions on text (Task 4) | 22 | 6.2727 | 2.47236 |
| Total | 76 | 7.0526 | 2.12850 |

Table 3: ANOVA of During Task Performance

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 25.115 | 3 | 8.372 | 1.915 | .135 |
| Within Groups | 314.675 | 72 | 4.370 | | |
| Total | 339.789 | 75 | | | |

Then, in order to measure and compare the participants vocabulary gain across the four tasks and also to explore the second research question, a one-way ANOVA was conducted using the posttest scores as the dependent variable and task types as the independent variable (Tables 4 and 5).

Table 4: Descriptive Statistics of the Participant Performance on the Posttest

| Tasks | N | Mean | Std. Deviation |
|---|---|---|---|
| Task 1 | 20 | 6.2500 | 2.35361 |
| Task 2 | 18 | 3.7778 | 2.55655 |
| Task 3 | 16 | 3.3125 | 2.46221 |
| Task 4 | 22 | 2.5000 | 1.58865 |
| Total | 76 | 3.9605 | 2.63156 |

As Table 4 shows, the highest mean score belongs to the group who performed Task 1 (i.e. reading plus fill-in-the-blank) followed by Task 2 (reading and sentence rewording) and then task 3 (reading and choosing definitions). The least mean score was acquired by the participants who performed task 4 (reading and multiple-choice on text). The results of the one-way ANOVA revealed a statistically significant difference among the four tasks (Table 5). A Tukey post hoc was thus conducted to locate where the difference lays (Table 6).

Table 5: ANOVA of the Participants Performance on the Posttest

|  | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 159.083 | 3 | 53.028 | 10.597 | .000 |
| Within Groups | 360.299 | 72 | 5.004 |  |  |
| Total | 519.382 | 75 |  |  |  |

The results demonstrated that the mean score of Task 1 (reading plus fill-in-the-blank) was significantly different from those of the other tasks. However, no significant difference was observed among tasks 2,3 and 4 (Table 6).

Finally, to check the results against the assumptions of the two frameworks, a cross-task comparison was made (Table 7). As Table 7 clearly shows, the assumptions made by the ILH were partially supported; however, those made by the TFA were strongly supported. Based on the assumptions made by both frameworks, Tasks 2, 3, and 4 had equal rankings (ILH= 3, TFA = 6) and accordingly they were expected to be equally effective in vocabulary acquisition. These assumptions were strongly supported in the study. However, the two frameworks differed in that the ILH considers Task 1 as the least effective task but TFA considers it as the most effective one. The findings were in line with the assumptions of TFA. Accordingly, the ILH is partially supported but the TFA is strongly supported by the study.

Table 6: Post-hoc Multiple Comparisons across the Four Tasks

| (I) Task | (J) Task | Mean Difference (I-J) | Std. Error | Sig. |
|---|---|---|---|---|
| Task 1 | Task 2 | 2.47222* | .72678 | .001 |
|  | Task 3 | 2.93750* | .75031 | .000 |
|  | Task 4 | 3.75000* | .69114 | .000 |
| Task 2 | Task 1 | -2.47222* | .72678 | .001 |
|  | Task 3 | .46528 | .76861 | .547 |
|  | Task 4 | 1.27778 | .71096 | .076 |
| Task 3 | Task 1 | -2.93750* | .75031 | .000 |
|  | Task 2 | -.46528 | .76861 | .547 |
|  | Task 4 | .81250 | .73500 | .273 |
| Task 4 | Task 1 | -3.75000* | .69114 | .000 |
|  | Task 2 | -1.27778 | .71096 | .076 |
|  | Task 3 | -.81250 | .73500 | .273 |

\* The mean difference is significant at the 0.05 level.

Table 7: 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 1> Task 2 | 2.47222* | X | √ |
| Task 1> Task 3 | 2.93750* | X | √ |
| Task 1> Task 4 | 3.75000* | X | √ |
| Task 2 > Task 3 | .46528 | √ | √ |
| Task 2 > Task 4 | 1.27778 | √ | √ |
| Task 3 > Task 4 | .81250 | √ | √ |

In order to examine and compare the predictive power of the TFA and ILH with more precision, a hierarchical multiple regression was conducted. So, based on the students' test scores, weighted scores for the ILH and TFA were first calculated for each task (Tables 8 and 9). weighted scores were calculated using the weight given to different components in each framework.

Table 8: Percentage of Distribution of the Components in TFA

Task 1 = 100/7 = 14.28 ≃14 [ 0-28 = 1; 29-57= 2; 58-72=4; 72-100=5]

Task 2 = 100/6 = 16.66 [0-16=1; 17-49= 2; 50-82=4; 82-100=5]

| Task Types | Motivation | Noticing | Retrieval | Generation | Retention |
|---|---|---|---|---|---|
| Task 1 (reading + fill-in-the-blank) | 2 (28%) | 2 (28%) | 0 | 1 (14%) | 2 (28%) |
| Task 2 (reading + sentence rewording) | 1 (16%) | 2 (32%) | 0 | 2 (32%) | 1 (16%) |
| Task 3 (reading + choosing definitions) | 2 (33%) | 2 (33%) | 1 (17%) | 0 | 1 (17%) |
| Task 4 (reading + multiple-choice on text) | 1 (17%) | 1(17%) | 2(32%) | 1(17%) | 1(17%) |

Task 3 = 100/6 = 16.66 ≃ 16 [0-32 = 1; 33-65=2; 66-82=3; 83-100=5]

Task 4 = 100/6 =16.66 [ 0-16 = 1; 17-33 = 2; 34-66 = 3; 67-83 = 4; 84-100 = 5]

Table 9: Percentage of Distribution of the Three Components in the ILH

| Tasks | Need | Search | Evaluation |
|---|---|---|---|
| Task 1 (reading + fill-in-the-blank) | 1(50%) | 0 | 1(50%) |
| Task 2 (reading + sentence rewording) | 1 (33%) | 0 | 2 (67%) |
| Task 3 (reading + choosing definitions) | 1 (33%) | 1(33%) | 1(33%) |
| Task 4 (reading + multiple-choice on text) | 1 (33%) | 1(33%) | 1 (33%) |

Task 1 = 100 /2 = 50 [0-50 = 1 51-100 = 3]

Task 2= 100/3 = 33.33 [0-33 = 1 34-100= 3]

Task 3 = 100/3 = 33.33 [ 0-33=1 34-67=2 68-100=3]

Task 4 = 100/3 = 33.33 [0-33 = 1; 34-67=2; 68-100=3]

Accordingly, the students' obtained scores were converted into percentile scores using the formula: n/10 x 100 = p% (n = scores of the task; 10 = the number of target words). Considering the different components proposed by the ILH and the TFA, the percentile scores were divided into percentile ranks. For example, for the TFA, the scores of each task were converted into percentile ranks based on the five components of motivation, noticing, retrieval, generation and retention (Table 8). Similarly, for the ILH, the scores of each task were converted into percentile rank based on the three components of need, search, and evaluation proposed for each task (Table 9).

The obtained percentile score could explain an index in the ILH and the TFA. For instance, the percentile score of a participant who did the first task and got a score of 6 in the posttest, is 60%. Based on the five components of the TFA, a percentile score of 60% falls between 58 and 72 percentile rank which is equivalent to the TFA index of 4. Similarly, based on the components of the ILH, a percentile score of 60%, falls between 51 and 100 which is equivalent to the ILH index of 3.

Then a hierarchical multiple regression was conducted to see which of the two frameworks could have a better explanatory power of the amount of vocabulary gain (Table 10). To this end, the obtained scores from pretest to posttest were considered as the dependent variable and the two ILH and TFA frameworks as the independent variables. The two independent variables were entered into the equation in different orders.

Table 10: Multiple Regression Analysis of Variables Predicting Word Gains

| City or Town | R | R2 | △R2 | F | df | Sig. F change |
|---|---|---|---|---|---|---|
| **Model 1** | | | | | | |
| TFA | .892 | .796 | .793 | 284.906 | (1,73) | .000 |
| ILH | .912 | .832 | .827 | 178.149 | (2,72) | .000 |
| **Model 2** | | | | | | |
| ILH | .821 | .675 | .670 | 151.276 | (1,73) | .000 |
| TFA | .912 | .832 | .827 | 178.149 | (2,72) | .000 |

In Model 1, the TFA was first entered and it accounted for 79% of variance in the amount of the gains which was significant and the on the next step, the ILH was entered which showed 82% significant difference from pretest to posttest (Table 10).

Conversely, in Model 2, first the ILH was entered into the equation followed by the TFA. This time the ILH and TFA respectively justified 67 % and 82% of the variance in the quality of gains which were both significant. The results of the multiple regression analysis revealed no significance difference between the two frameworks in terms of their explanatory power in accounting for vocabulary gains.

## 5. Discussion

This study examined to what extent vocabulary tasks with similar and different indexes given by the ILH and TFA contribute to L2 vocabulary learning. To this end, students' performance both during the task and on the posttest were examined. During task performance was measured based on the participants' correct responses to the activities following the reading passage in each task or the number of correct answers in the fill-in-the-blank task. In line with the study conducted by Jafari Gohar et al. (2018), the results of ANOVA for during task performance revealed no significant difference among the tasks and thus the assumptions made by the TFA and ILH were not supported.

Given that none of the words was familiar or known by the participants, the means across the four tasks in the posttest suggested that all four task types facilitated vocabulary learning. The results obtained by analyzing the posttest scores revealed a significant difference between task 1 (ILH = 2, TFA=7) and the other three tasks (ILH=3, TFA= 6) providing a strong support for the technique feature analysis. Moreover, in line with the findings reported by Hu and Nassaji (2016) and Jafari Gohar et al. (2018), the findings of the present study showed that the task scored higher by the TFA (i.e. reading and fill-in-the-blank) resulted in better vocabulary acquisition than other three tasks; however, unlike the study conducted by Hu and Nassaji (2016), in this study, the results of the hierarchical regression did not confirm the TFA to be a better predictor of vocabulary task effectiveness. Nonetheless, considering the results of ANOVA for the posttest, it could generally be inferred that the TFA was more satisfactory.

The effectiveness of reading and fill-in-the-blank task can be explored from different aspects. According the TFA, fill-in-the-blank task has an index of 7 because (1) it involves a clear vocabulary learning goal: the participants are required to match the target words with appropriate contexts; (2) a meaningful context with semantic associations can motivate learning; (3) it raises awareness of new vocabulary learning and focuses learners' attention on the target words semantically and syntactically. To fill in the blanks, learners need to understand the meaning of words and their association with the surrounding context (the involvement load hypothesis, which is based on the depth of processing hypothesis, associates more semantic processing with deeper processing and ignores syntactic processing (Al-Had Laq, 2003)). However, a context also contains linguistic context in which the grammatical function of the target word and its surrounding words or phrases are also taken into account (Al-Had Laq, 2003); (4)it involves receptive generative use of the target words because learners need to compare different words to select the most appropriate one for a given context; (5) successful linking of form and meaning is ensured when the target words are glossed and (6) no inference is involved. The fill-in-the-blank task is especially different from the other tasks in terms of retention which ensures successful form-meaning relation which may have more to say in vocabulary learning and play an important role in vocabulary learning tasks and in this study this feature was achieved through glossing.

In the current study the target words had been glossed. One of the reasons that the ILH index of fill-in-the-blank activity in this study is 2 is that it has been glossed. Whenever a task is glossed, according to the ILH, its search component index is zero. In fact the value and effect of glossing on vocabulary acquisition has been completely ignored by the hypothesis. However, it has indirectly been taken into account by the TFA under the component of retention as an attribute which consolidates successful form-meaning link and also under the component of noticing since it attracts learners' attention to the target words.

Glossing can lead to successful acquisition of form-meaning link which is the first and most essential lexical aspect to be acquired (Schmitt, 2008). In addition, since the languages have conceptually a lot in common (Swan, 1997, as cited in Joyce, 2015), the use of L1 in L2 learning can provide a shortcut to vocabulary acquisition (Scott & De la Fuente, 2008). Moreover, it can provide additional exposure to the target words (Joyce, 2015).

Hulstijn and Laufer (2001) deem fill-in-the-blank exercises as a superficial or passive use of the vocabulary. However, as Folse (2006) properly explains, not only does this activity involve deep processing but also it is highly efficient in terms of student and teacher time required. "When a learner encounters a blank in a sentence, in a vocabulary exercise, however, who can say that the learner's process in trying out the various words in this slot, perhaps by translating many of the words or perhaps by remembering tidbits about some of the words…is not indeed deep processing of or high involvement with the word?" (Folse, 2006; p. 287).

**Chabahar Maritime University**

**Iranian Journal of English for Academic Purposes**                    ISSN: 2476-3187
IJEAP, (2017) Vol. 6 Issue. 2                    (Previously Published under the Title: Maritime English Journal)

## 6. Conclusion

Overall the results suggest that of the two frameworks, namely the TFA and the ILH, the former was a better predictor of lexical gains than the latter. It was evidenced by the findings that the task with the higher TFA score (fill-in-the-blank task) led to significantly better word retention in the posttest.

The findings of this study might provide a good foothold for language instructors and educators in selecting and designing vocabulary learning tasks. When designing a task, they can use TFA as a helpful framework against which the features of a task can be checked; and they can also prepare tasks which provide more learner engagement due to having more of the TFA features especially the generative component which can help learners to notice the gap of knowledge (Swain, 2005) and make them retrieve and rehearse the target words which in turn consolidate vocabulary knowledge (Keating, 2008; Zou, 2016; Laufer, 2006). Generation can provide learners with an opportunity to remember and in effect highlight the form-meaning relationship of vocabularies in their mind (Keating, 2008; Laufer, 2006). Activities which lead to successful linking of form and meaning through instantiation, imaging, glossing, etc., like fill-in-the-blank task in this study, might also be really effective in vocabulary acquisition. EFL teachers may sometimes integrate various tasks to use the merits of each task for vocabulary learning.

Furthermore, the results of the current study might provide useful insights for material developers and syllabus designers in their selection of effective vocabulary learning tasks. The framework can also be used to evaluate vocabulary activities in textbooks so that they can be modified in a way which triggers better vocabulary acquisition.

The current investigation like any other research in SLA is liable to some 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 Time on task was not considered in this study. Longer time on task or longer exposure to the target words might be an attribute which can affect vocabulary acquisition.

The study used the VKS in which students had to recall the meaning of the target words. The research findings might be different if a recognition test had been used. Research should be designed in which the posttest includes both recall and recognition so that the effect of task types on both recognizing and recalling words could be measured and compared. Moreover, the present study investigated the short-term effect of tasks on the retention of the target words. No delayed posttest was given to measure the long-term retention of the target words.

The tasks used in this study had close involvement load indexes and technique feature scores. The extent to which the components of these frameworks can contribute to vocabulary learning cannot be exactly measured.

Another limitation of the study is concerned with the number of participants which was not big enough. Future studies are needed with larger number of participants to compare and contrast these two features more delicately to better examine the predictability power of these two frameworks.

**References**

Al-Hadlaq, M.S. (2003). *Retention of words learned incidentally by Saudi EFL learners through working on vocabulary learning tasks constructed to activate varying depths of processing*. (Unpublished doctoral dissertation). Ball State University, India.

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

Cain, K., & J. Oakhill. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*, 431–443. doi:10.1177/0022219411410042.

Coady, J. & T. Huckin (Eds.). (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. New York: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. https://doi.org/10.2307/3587951.

Craik, F. & R. Lockhart.(1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671- 684.

Craik, F. & E. Tulving. (1975). Depth of processing and the retention of words in Episodic memory. *Journal of Experimental Psychology: General, 104* (3), 268-294. http://psycnet.apa.org/doi/10.1037/0096-3445.104.3.268.

Decarrico, J. (2001). Vocabulary learning and teaching. In Celce-Murcia, M. (ed.), *Teaching English as a second or foreign language*. Boston: Heinle and Heinle.

Ellis, R.(1994). *The study of second language acquisition*. New York: Oxford University Press.

Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*, 273–293. https://doi.org/10.2307/40264523.

Ghapanchi, Z., M. Davoudi & Z. Eskandari. (2012). Vocabulary learning through input and output tasks: investigating the involvement load hypothesis. *California Linguistic Notes, 37* (1), 1-18.

Hu, M. & H. Nassaji. (2012). Ease of inferencing, learner inferential strategies, and their relationship with retention of word meanings inferred from context. *Canadian Modern Language Review, 68*, 54-77.

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

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

Jahangiri, K. & I. Abilipour. (2014). Effects of collaboration and exercise type on incidental vocabulary learning: Evidence against involvement load hypothesis. *Procedia - Social and Behavioral Sciences*, *98*, 704 –712. https://doi.org/10.1016/j.sbspro.2014.03.471.

Jafari Gohar, M., M. Rahmanian, H. Soleimani. (2018). Technique feature analysis or involvement load hypothesis: estimating their predictive power in vocabulary learning. *Journal of psycholinguistic research, 47*(4),859-869. doi: 10.1007/s10936-018-9568-5.

Joyce, P. (2015). L2 vocabulary learning and testing: The use of L1 translation versus L2 definition. *The Language Learning Journal*, 1-12. http://dx.doi.org/10.1080/09571736.2015.1028088. (accessed 18 May 2015).

Karalik, T., & A. Merç. (2016). *The effect of task-induced involvement load on incidental vocabulary acquisition among ELT* students. Turkey: Anadolu University Graduate School of Educational Sciences MA thesis.

Keating, G. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research 12*(3), 365-386. https://doi.org/10.1177%2F1362168808089922.

Kim, Y. (2008). The role of task induced involvement load and learning proficiency in L2 vocabulary acquisition. *Language learning 58* (2), 285-325. https://doi.org/10.1111/j.1467-9922.2011.00644.x.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition: A rationale for pedagogy*, 20-33. New York: Cambridge University Press.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different. *Applied Linguistics*, *19*(2), 255-271. https://doi.org/10.1093/applin/19.2.255.

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

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

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

Lee, C. (2003). *The effect of involvement load on English vocabulary acquisition in an incidental learning situation*. (MA thesis). National Tsinghua University, Hsinchu, Taiwan.

Maftoon, P. & M. Sharifi Haratmeh. (2012). The relative effectiveness of input and output-oriented tasks with different involvement loads on the receptive and productive vocabulary knowledge of Iranian EFL learners. *The Journal of Teaching Language Skills (JTLS)*, 4 (2), 27-52. http://dx.doi.org/10.22099/jtls.2012.497.

Martínez-Fernández, Ana. (2008). Revisiting the involvement load hypothesis: awareness, type of task and type of item. In *Selected Proceedings of the 2007 Second Language Research Forum*, ed. Melissa Bowles, Rebecca Foote, Silvia Perpiñán, and Rakesh Bhatt, 210-228. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1746.

Mobarge, M. (1997). Acquiring, teaching and testing vocabulary. *International Journal of Applied Linguistics*, *7*(2), 201-222. https://doi.org/10.1111/j.1473-4192.1997.tb00115.x.

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, *37*(4), 645-670. https://doi.org/10.2307/3588216.

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review*, *61* (1), 107-134. https://doi.org/10.1111/j.1540-4781.2006.00431.x.

Nation, I. S. P. (1984). Vocabulary lists: Words, affixes and suffixes. *ELT Occasional Publication* 12. Wellington: Victoria University of wellington.

Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House Harper Row.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* New York: Cambridge University Press.

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

Paribakht, T. & M. Wesche. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise types. *The Canadian Modern Language Review*, *52*(2), 155-178.

Paribakht, T. S. & M. Wesche. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition,* 174-200. New York: Cambridge University Press.

Peters, E. (2012). The differential effects of two vocabulary instruction methods on EFL word learning: A study into task effectiveness. *IRAL*, *50*, 213–238. doi 10.1515/iral-2012-0009.

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. https://doi.org/10.1017/S0272263109090020.

Reynolds, B. (2014). Evidence for the task-induced involvement construct in incidental vocabulary acquisition through digital gaming. *The Language Learning Journal*, *45* (4), 466-484. http://dx.doi.org/10.1080/09571736.2014.938243. (accessed 26 Aug 2014).

Richeck, M. A. (1993). *World of words: Vocabulary for college students*, 4th edn. Boston: Houghton Mifflin.

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, *17*(2), 95-124.

Sarbazi, M. (2014). Involvement load hypothesis: Recalling unfamiliar words meaning by adults across genders. *Procedia - Social and Behavioral Sciences*, *98*,1686-1692.

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

**Chabahar Maritime University**

**Iranian Journal of English for Academic Purposes**                    ISSN: 2476-3187
IJEAP, (2017) Vol. 6 Issue. 2                    (Previously Published under the Title: Maritime English Journal)

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12* (3), 329-363.

Scott, V.M. & M.J. De la Fuente. (2008). What is the problem? L2 learners' use of the L1 during consciousness-raising form-focused tasks. *The Modern Language Journal*, *92*(1), 100–13.

Silva, B. & A. Otwinoska. (2017). Vocabulary acquisition and young learners: Different tasks, similar involvement loads. *IRAL*, 1-25. doi 10.1515/iral-2016-0097. (accessed 13 May 2017).

Soleimani, H. & M. Rahmanian. (2014). The role of language glossing in a rooted theory: The involvement load hypothesis. *International Journal of Applied Linguistics and English Literature 3*(4), 6-13. doi: 10.7575/aiac.ijalel. v.3n.4p.6.

Soleimani, H., M. Rahmanian & K. Sajedi. (2015). A revisit to vocabulary acquisition in involvement load hypothesis. *Procedia - Social and Behavioral Sciences*, *192*. 388-397. https://doi.org/10.1016/j.sbspro.2015.06.055.

Soleimani, H. & A. Rostami Abu Saeedi. (2015). The interaction between involvement load hypothesis evaluation criterion and language proficiency: a case in vocabulary retention. *Iranian Journal of Applied Language Studies*, *8*(1), 173-194. http://dx.doi.org/10.22111/ijals.2016.3025.

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

Swan, M. (1997). The influence of the mother tongue on second language acquisition and use. In N. Schmitt & M. McCarthy (eds.), *Vocabulary: Description, acquisition and pedagogy* ,156-180. Cambridge, UK: Cambridge University Press.

Tahmasbi, M. & M. Farvardin. (2017). Probing the effects of task types on EFL learners' receptive and productive vocabulary knowledge: The case of involvement load hypothesis. *Sage Open*, *7*(3). journals.sagepub.com/home/sgo,1-10. doi: 10.1177/2158244017730596.

Yaqubi, B., R. Rayati, & N. Allemzade Gorgi. (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 (JTLS)*, *2* (1),145-163. *http://dx.doi.org/10.22099/jtls.2012.404.*

Zou, D. (2016a). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research*, *21*(1), 54 -75. doi: 10.1177/1362168816652418.

Zou, D. (2016b). Comparing dictionary-induced vocabulary learning and inferencing in the context of reading. *Lexikos*, *26*, 372-390. http://dx.doi.org/10.5788/26-1-1345.