Effects of Input and Output-oriented Tasks with Different Involvement Loads on the Receptive Vocabulary Knowledge of Iranian EFL Learners

Parviz Maftoon

Professor, Islamic Azad University, Science and Research Branch

pmaftoon@gmail.com

Marzieh Sharif Haratmeh*

Ph.D. candidate, Islamic Azad University, Science and Research Branch

msharifh2000@yahoo.com

Abstract

Framed in a cognitive approach to task-supported L2 vocabulary learning, the present study used a pedagogical approach to investigate the relative effectiveness of tasks with different involvement loads on the vocabulary knowledge of Iranian EFL learners. The goal was to investigate the way that the construct of involvement load is related to the Input Hypothesis (Krashen, 1985) and the Output Hypothesis (Swain, 1996) to see whether the involvement load or input/output-orientation of tasks is the determining factor in task effectiveness. A quasi-experimental design with a pretest-treatment-posttest sequence was used in this study. The participants were 127 university students from four General English classes at Islamic Azad University, Mobarakeh Branch, who were assigned to four instructional groups. Contrary to the predictions of the Involvement Load Hypothesis (Laufer & Hulstijn, 2001), the results of the study indicated that involvement load is not the only determining factor in task effectiveness, but input/output-orientation of tasks is also a decisive parameter in task effectiveness. While Laufer and Hulstijn's proposal is a valuable first step towards building a theory of vocabulary learning, the results of the study indicated that involvement index may well not function independently of the task type, i.e., input or output orientation of a word-focused task.

***Keywords:* EFL, Involvement load, Input-oriented tasks, Output-oriented tasks, Vocabulary learning, and Word-focused tasks.**

Introduction

Language learning is probably the most complex type of learning which human beings need to accomplish. Within this complex task, vocabulary, as one of the important components of language knowledge seems to be of critical importance to both native and non-native speakers of a language. Whereas the number of syntactic patterns that language learners have to learn is finite, vocabulary is an unending task which continues to be learned throughout one's lifetime. Vocabulary knowledge is considered a prerequisite

---

factor in reading comprehension, a fact that has been recognized in the literature available so far (e.g., Decarrico, 2001; Garcia, 1991; Hazenberg & Hulstijn, 1996; Laufer, 1992; Nation, 1993; Paribakht Wesche, 1997; Wesche & Paribakht, 1998). Research has shown that lexical errors impede communication more seriously than grammatical ones (Ellis, 1994; Laufer, 1998).

Most research in instructed Second Language Acquisition (SLA) is conducted under the assumption that some interventions are better than others. While many researchers have tended to agree that the vast majority of vocabulary is learned indirectly through reading and listening (e.g., Nagy, Herman, & Anderson, 1985), others (e.g., Laufer, 2005; Nation, 2001) indicate that direct learning tasks, such as word pairs, are effective methods of quickly acquiring the vocabularies of a language. Laufer (2002), who has surveyed a large body of evidence, reports that vocabulary gains are greater when learners engage in word-focused tasks. One essential question is then whether instructional tasks can be classifed in terms of their vocabulary learning effectiveness. Language teachers and learners would like to know the ways that instructional programs might foster the acquisition of new words. Of primary concern to researchers and language teachers in the feld of L2 vocabulary learning is to identify those tasks that provide the best opportu- nity for learners to notice and elaborate on new words.

Much of the current research on vocabulary acquisition is implicitly or ex-plicitly based on a cognitive processing view of learning. Of all that have been done in this feld, the Involvement Load Hypothesis (ILH) (Laufer & Hulstijn, with the basic contention that retention of unfamiliar words is generally conditional upon the degree of involvement in processing theses words draws the researchers' attention. Involvement is operationalized by tasks designed to vary in the degree of need, search, and evaluation. The need component is the motivational, non-cognitive dimension of involvement which can be in moderate and strong degrees. Laufer and Hulstijn interpret this notion based on a drive to comply with the task requirements, whereby the task requirements can be either externally imposed or self-imposed. According to Laufer and Hulstijn, need is moderate when it is imposed by an external agent. An example is the need to use a word in a sentence that the teacher has asked for. Need is strong when it is intrinsi- cally motivated, that is, self-imposed by the learner. An example is the decision to look up a word in an L1-L2 dictionary when writing a composition.

Search and evaluation are the two cognitive components of involvement, contingent upon allocating attention to form-meaning relationship (Schmidt, 1990). Search is the attempt to find the meaning of an unknown L2 word by consulting a dictionary or another authority. Evaluation entails a comparison of a given word with other words, a specific meaning of a word with its other meanings, or comparing the word with other words in order to assess whether a word does or does not ft its context. The kind of evaluation that entails recognizing differences between words (as in a fill-in task with words provided), or differences between several senses of a word in a given context, is referred to as moderate. Evaluation that requires a decision such as how additional words will combine with the new word in a sentence or context is referred to as strong evaluation (Laufer & Hulstijn, 2001).

Each of the above three factors can be absent or present when processing a word in a natural or artifcially designed task. The combination of factors with their degrees of prominence constitutes involvement load. The basic proposition of the ILH is that

retention of unfamiliar words is conditional, in general, upon the degree of involvement in processing the unknown words. In other words, it is conditional upon who has set the task and whether it has to be compared or combined with other words. The greater the involvement load, the better the retention. One of the major areas of interest is the way that the construct of involvement load is related to the Input Hypothesis (Krashen, 1985), and the Output Hypothesis (Swain, 1996). According to Laufer and Hulstijn (2001), the hypothesis does not predict that any output task will lead to better results than any input task. It predicts that higher involvement in a word induced by the task will result in better retention regardless of whether it is an input or an output task. This motivated the researcher to test this contention. Accordingly, a particularly interesting comparison would involve conditions where the input and output tasks have identical involvement loads. In other words, if involvement load is the determining factor in task effectiveness, irrespective of whether the task is input oriented or output oriented, the two conditions should yield similar retention results.

Following the ILH proposal, many investigations have tried to examine the accuracy of the claim made by Laufer and Hulstijn (2001). Some studies (Fuente, Hulstijn & Laufer, 2001; Joe, 1995; Keating, 2008) support the claim that tasks with higher involvement loads generally lead to greater gains in short-term and, in some cases, long-term word retention. However, contra the predictions of the ILH, Folse (2006) and Kim (2008) found that word-focused tasks with higher involvement loads was as effective as those with lower involvement loads.

As mentioned earlier, studies measuring the relative effectiveness of word-focused tasks with different involvement loads on learners' vocabulary knowledge have produced conficting findings. Given the mixed results, it seems necessary to conduct more studies before rushing to support the ILH claims. This motivated the researcher to set out the present study with the intent to determine if the ILH could be generalized to other EFL contexts, in this case Iran. The main goal of the study was to investigate the effect of tasks with different involvement loads on the vocabulary knowledge of EFL university students. Furthermore, in this study an attempt has been made to relate the construct of involvement load to the Input and Output Hypotheses. The goal was to make a comparison between input and output-oriented tasks with identical involvement loads to see which one plays a more determining role in task effectiveness.

2. Method
2.1. Participants
Primarily, a group of 159 EFL students, who were all Persian native speakers, from four intact university-level classes at Islamic Azad University, Mobarakeh Branch, participated in this study. They were B.A. and B.S. students who had en-rolled in General English course as a required part of the university curriculum. As the participants were assigned to classes by the university registration offce, it was practically impossible to disrupt the regular schedules. Thus, in order to re- duce the effects of extraneous variables and selection bias, the four intact classes were randomly assigned to four treatment groups.

Out of the original pool, only the data from 127 students entered into fnal data analyses. This was due to the elimination of 32 subjects who did not meet the established criteria of the study. Three main considerations were taken into ac- count to select the data for the final analyses. The first criterion was the language proficiency level of the participants measured by PET. The second criterion was the participants'

attendance in all sessions. The third criterion was the lack of exposure to the target words outside the class during the treatment period. This left a group of 127 participants, both male and female, in four study groups. The age of the participants ranged from 18 to 27. In this study neither gender nor age was a variable.

2.2. Materials

The materials used in this study were four word-focused tasks (two input-based and two output-based) with different involvement loads based on the ones used in the related literature (Hulstijn & Laufer, 2001; Laufer, 2005; Laufer& Hulstijn, Webb, 2002), and a translation test for measuring the receptive knowledge of meaning.

2.2.1. Word-focused Tasks
2.2.1.1. Input Tasks with an Involvement Load of 2

The participants assigned to Task 1 were provided with sentences containing the target words. The target words were highlighted in bold print to help the par-ticipants notice the words, and were glossed in the L1 (Persian) in the margin of the sentences. Each word had more than a single meaning. The task induced mod-erate need (1) because the participants required the meaning of the target word for understanding the sentence, no search (0) as the participants did not consult a dictionary or other sources to fnd the meaning of the target word, and moderate evaluation (1) since it required the participants to recognize differences between several senses of a word and choose the one that best fitted the context. So the Involvement Load Index (ILI) for this task was 2 (1+0+1).

2.2.1.2. Input Task with an Involvement Load of 3

The participants assigned to Task 2 were provided with the same sentences and the same target words as participants who received Task 1. The target words were highlighted in bold print to be noticed, and the participants' task was to look up the targets word in a dictionary. To control the variable, and use the same dictionaries, the participants took part in computer-laboratory classes and used the same software dictionary, Babylon. The tasks induced moderate need (1), moderate search (1), since the meaning of the word was looked up, and moderate evaluation, so the ILI for this task was 3 (1+1+1).

2.2.1.3. Output Task with an Involvement Load of 2

The participants assigned to Task 3 were given the same sentences and the same target words. For this group, however, the bold-faced target words were de- leted from the sentences and replaced with blank spaces. The target words, along with some additional words not present in the original sentences, were printed in alphabetical order as a list on a separate page with their L1 translations. The task required the participants to read the sentences and fll in the gaps with target words from the list. In terms of involvement load, the task induced moderate need (1), no search (0), and moderate evaluation (1). Its ILI was 2 (1+0+1).

2.2.1.4. Output task with an Involvement Load of 3

The participants assigned to Task 4 were given strings of words including the target word and were asked to rearrange them into meaningful and grammatically correct sentences. In terms of involvement load, this task induced moderate need (1), no search (0), and evaluation was strong (2) because it required the partici- pants to decide on how the additional words were combined with the target word in a sentence. So, the ILI was 3(1+0+2).

2.2.2. Test of Receptive Knowledge of Meaning

In this study, the receptive knowledge of meaning was measured using a receptive translation test. On this test, a list of 33

target words was given to the participants with a blank beside each word. The participants' task was to write the Persian equivalent of the target words to score a correct response. Nation (2001) suggests that having learners write the translation of words may be a more effective method of measuring meaning than multiple-choice tests. Furthermore, trans- lation tests reduced the time needed for students to complete the test.

2.3. Procedure

The general procedures employed in this study can be divided into two main phases: developing and piloting the materials, and conducting the main phase of the study.

2.3.1. Phase one: Developing and Piloting the Materials

2.3.1.1. Target words

Primarily, a group of 63 low frequency words from 3,000 Word Frequency List (Nation, 2001) was selected for the study. The target words were unfamiliar to the participants as they were checked in the pretest. Factors considered when establishing the number of the target words were subject fatigue, time needed to complete the tasks and tests, and the results of item characteristics in the pilot study. Many of the words were deleted from the list because of unacceptable item characteristics Indices. It left a group of 33 words as the target words in the study. The words included Nouns (11 words), Verbs (11 words) and Adjectives (11 words). The justifcation was that nouns, verbs, and adjectives were the most common parts of speech as grammatical functions of the words. Only one sense of the target words appeared in the sentences. The same target words were used in the treatment tasks and the assessment.

2.3.1.2. Sentence Contexts

The target words were presented in short sentences. The sentences were selected from Oxford Learner's Dictionary and British National Corpus. Three factors were considered when selecting the sentences: the number of words, the frequency of the words, and the ease in comprehending the sentences. The average length of the sentences for the target words was 10.8 words. Sentences that contained words unknown to the subjects in the pilot study were eliminated.

2.3.1.3. Developing and Piloting Word-focused Tasks

The tasks were developed by the researcher after reviewing the relevant literature and were based on the tasks used in the published research. Before the com-mencement of the main study, the tasks were piloted on a sample of 110 students who had enrolled in General English courses in the first semester. All teaching procedures were trialed at this stage. The piloting was done to determine the time required for doing the tasks in the main study, to check the practicality of using the tasks in the class, and to determine the possible practical problems in imple-menting the tasks. Time on task was different for all four tasks. Tasks 1, 2, 3, 4 took about 35-40, 60-70, 40-45, and 70-80 minutes, respectively. Task 4 was the most demanding and time consuming. It is often argued that time on task should be kept identical in research on task effectiveness. Yet, in this study time on task was regarded as an inherent property of the task, not as a separate variable.

2.3.1.4. Developing and Piloting the Vocabulary Test

The vocabulary test used in this study was also developed by the researcher based on the measurements available in the relevant literature (e.g., Fuente, 2006; Hulstijn & Laufer, 2001; Keating, 2008; Webb, 2002). The items were aimed at assessing the participants' receptive vocabulary knowledge of meaning. The number of items was far more than what was really needed in the main study. The goal of piloting the test

was to examine item characteristics and test characteristics. The results of item characteristics analyses (Item Facility and Item Discrimina- tion) left 36 items for the main study. To ensure the content validity of the test, the content of the test and the content of the tasks were examined carefully and criti- cally by expert judges. As a result of the panel views, several items were either discarded or modifed. Furthermore, an internal-consistency method (Cranach's Alpha) was used to calculate the reliability of the test. The reliability was estimat- ed from the test scores in the two posttests. The underlying reason was that there was very little variability in the pretest scores which would decrease the reliability indices. The reliability measures for the immediate posttest (.79) and the delayed posttest (.76) were high enough to confirm the reliability of the test. As Jafarpoor (1992) states, for teacher-made tests, a reliability of .60 and above is adequate.

2.3.2. Phase Two: Main Study

The main study was conducted in the semester following the pilot semester. The primary objective of General English courses is to develop the learners' reading comprehension abilities. The students should be taught a reading book recommended by the English Department of the University. Each chapter in the book includes a list of new words, a passage for comprehension, some text-related comprehension questions, and grammar.

In this study, training the groups and administration of the measurements were performed by the researcher, who was also the instructor, to control for teacher variable and ensure the elimination of possible differences in the implementation of the instructional tasks by different teachers. This could help avoid the possible threats to the internal validity of the study. The subsequent steps taken in the main phase of the study were as follows:

1. Administering PET: The PET test was administered in the second week of the semester in order to ensure the homogeneity of the four groups in terms of their language proficiency before the commencement of the study. To make sure that the participants were homogenous in terms of their language proficiency level, reading and writing sections of the Preliminary English Test (PET) were adminis- tered to the original pool before the commencement of the study (the listening and speaking sections of the test were not administered due to practical limitations). The mean (M= 21.35) and standard deviation (SD = 7.14) of the participants scores on the PET test were used as the criteria for choosing the participants. The participants who scored higher or lower than one standard deviation plus/minus the mean of the sample (M+/- SD) were considered as high and low proficient, respectively, and were excluded from the final data analyses. The data belonging to those participants who scored one standard deviation above or below the mean were included in the final data analyses. The total scores of the remaining subjects were then entered into one-way ANOVA with the alpha set at p<.05. The results revealed no significant difference across the groups as far as language proficiency level was concerned: F (3,136) = 0.961, p = .394> .05.

2. Administering the Pretest: The pretest was administered in the third week of the semester. The first purpose of administering the pretests was to examine the likelihood of the target-word familiarity among the participants before the treatment commencement. The goal was to ascertain the knowledge of the target words and to select a group of target words unknown to all participants. Three participants were excluded after the test, all other participants showed no previous knowledge of the target words. The second purpose was to

address the research questions concerning the possible effectiveness or superiority of one word-focused task over the other in developing the participant's receptive vocabulary knowledge of meaning.

3. Training the treatment groups: Since the unfamiliarity of the participants with the word-focused tasks could affect the tasks, the groups were first given the op- portunity to get familiar with the tasks and practice the kind of activities they were supposed to receive. For three weeks they practiced the list of the new words of their regular course book in the way they were supposed to practice the target words of the study. It was effective especially for the second (looking up the words in a software dictionary) and the fourth (sentence making) groups.

4. Implementing the treatment: The instructional treatments in the groups started in the seventh week and within a three-week interval of the pretest administration. All instruction and assessment took place in the participants' regular class time by the instructor. All participants were informed that they were participating in a study. They were told that the research involved the relationship between vocabulary learning tasks and their effects on vocabulary learning. According to Hulstijn (2003), methodologically speaking, test announcement is part of intentional vocabulary learning.

The participants took all tasks and posttests within regular classrooms during regular class time. On the treatment day, the participants in each of the four classes were given one of the experimental tasks. The tasks were completed by individual learners. As word-focused tasks will be of little use without providing the students with the feedback on the accuracy of their choices or activities, the instructor pro- vided any additional information. Whenever they needed, the participant could enjoy peer feedback while completing the tasks. Cognizant of the fact that time on task would vary among the treatment groups, the researcher asked participants to turn their worksheets face down when completed and waited for others to finish. In all the groups, the worksheets were collected after the completion of the tasks.

5. Administering the Immediate Posttest (IP): To assess initial recall of the target words, the IP was administered to the groups after the completion of the tasks. The same test in pretest was used as posttest.

6. Administering the Delayed Posttest (DP): The DP was administered to the groups to examine whether benefts of each of word-focused tasks would last over time or not. The DP was administered after a one-month interval. During the interval, the participants continued working on the reading comprehension book with no contact with the target words.

An attempt was made to equalize the administration procedures in all groups to increase the reliability and the internal validity of the measures (Arey, Jacobs, & Razavieh, 1996). In all test administrations, the instructor was present for clarifying the ambiguities for the examinees. After the required data were collected, they were subjected to different quantitative analyses.

3. Results

In order to investigate whether each individual study group (Groups 1, 2, 3, and improved in the receptive knowledge of meaning from the pretest to the IP and the DP, the within group changes over time were analyzed both descriptively and inferentially. It is depicted from the means and standard deviations of the groups' raw scores on the receptive test of meaning over the three test administration (Table 1) that all the study groups improved from the pretest to the IP. Although there was a decline on the DP for all the study groups, none of the groups returned to the same level of performance it was before the task-oriented instruction, showing the effectiveness of the word-focused tasks on the participants' receptive vocabulary knowledge of meaning over time.

Table 1.*Descriptive Statistics for the Study Groups (Receptive Knowledge of Meaning)*

		Pretest		Immediate Posttest		Delayed Posttest	
Groups	N	M	SD	M	SD	M	SD
1	35	.00	.00	13.09	4.61	6.09	2.93
2	26	.00	.00	13.08	3.46	6.27	2.09
3	31	.00	.00	14.90	4.84	6.81	1.94
4	35	.00	.00	15.74	5.22	7.80	3.58
Total	127						

Note: The maximum score for each test is 33

Figure 1 graphically shows the difference between the participants' receptive knowledge of meaning in different study groups. As it shows, there is an improvement in the receptive vocabulary knowledge of meaning from the pretest to the IP in all the study groups. However, there is a decline on the DP for all the study groups.
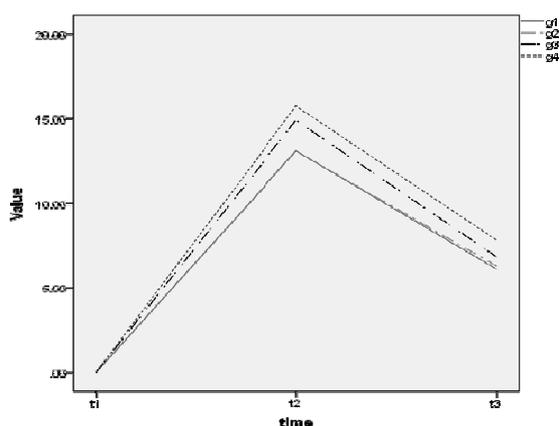

Figure 1. Groups' Performance on the Receptive Test of Meaning over Time

Although the results of within-group comparisons indicated that the word-focused tasks were all effective in improving the receptive vocabulary knowledge of meaning in each single group, they could not show if the results were statistically equal for all the groups or not. Thus, a one-way ANOVA was conducted to examine simple main effects of word-focused tasks on the IP. The results of ANOVA conducted on the receptive meaning scores of the groups on the IP revealed a significant difference in how the study groups performed on the test of receptivemeaning ($F(3,123)=2.695$, $p=004<.05$).

The results of ANOVA (Table 2) indicated that word-focused tasks did not produce equal results in the groups' receptive vocabulary knowledge of meaning. In order to specifically locate the differences among the study groups, LSD, as an adjustment for pair-wise comparisons was used to detect the precise location of the differences. The results indicated that there were significant differences be- tween Groups 1 and 4, and Groups 2 and 4. In other words, Group 4 outperformed Groups 1 and 2 in the receptive measure of meaning. The results indicated no significant differences between Groups 1 and 2, Groups 1 and 3, Groups 2 and 3, and Groups 3 and 4. The results of post hoc analyses are summarized as follows: Group 4>Group 1=Group 2=Group 3; Group 3=Group 4 (with> meaning better than, and = indicating no significant differences).

To see whether there were any significant differences among the study groups on the DP another one-way ANOVA was conducted to examine the main effects of word-focused tasks on the DP. The results of ANOVA conducted on the receptive meaning scores of the study groups on the DP revealed a significant difference in how the study groups performed on the test of receptive meaning $(F(3, 123)= 2.592$, $p= .005<.05)$.

Table 2.*Results of ANOVA on IP (Receptive Knowledge of Meaning)*

	SS	DF	DM	F	Sig.
Between Groups	174.441	3	58.147	2.695	.0049
Within Groups	2653.984	123	21.577		
Total	2828.425	126			

Table 3.*Results of ANOVA on IP (Receptive Knowledge of Meaning)*

	SS	DF	DM	F	Sig.
Between Groups	60.081	3	20.027	2.592	.005
Within Groups	950.297	123	7.726		
Total	1010.378	126			

The results of ANOVA (Table 3) indicated that the word-focused tasks did not have the same effect on the study groups' receptive vocabulary knowledge on the DP. In order to specifcally locate the differences among the study groups, LSD tests were run on the data. The results of pair-wise comparison indicated that there were not signifcant differences between Groups 1 and 2, Groups 1 and 3, as well as Groups 2 and 3, and Groups 3 and 4. However, a signifcant difference was found between Groups 1 and 4, as well as Groups 2 and 4.The results can besummarized as follows: Group 4>Group 1=Group 2; Group 4=Group 3; Group 3=Group 1=Group 2.

The main goal of the study was to investigate the differences between Group with an Involvement Load of 2 and Group 2 with an involvement load of 3 (both input-oriented tasks) on one hand, and Group 3 with an Involvement Load of 2, and Group 4 with an Involvement Load of 3 (both output-oriented tasks), on the other hand. The results of pair-wise comparisons indicated that the differences between Group 1 and Group 2 were not significant on the IP and on the DP measuring the receptive knowledge of meaning. On the other hand, the results of pair-wise comparisons between Group 3 with an involvement load of 2 and Group 4 with an Involvement Load of 3 (both output-oriented tasks) indicated no significant differences on the IP and DP.

Furthermore, the study aimed at investigating the significance of differences among the study groups as far as the input/output-orientation of the tasks was concerned. The goal was to investigate the differences between Group 1 and Group 3, with the involvement load of 2, and Group 2 and Group 4, with the involvement load of 3. The results of pair-wise comparisons indicated that there were no sig- nifcant differences between Group 1 and Group 3 neither on the IP nor on the DP measuring the receptive vocabulary knowledge of meaning. On the other hand, the results of pair-wise comparisons between Group 2 and Group 4 (both with the involvement load of 3) indicated significant differences between these two groups' receptive vocabulary knowledge of meaning not only on the IP but also on the DP. In other words, although the involvement load in both groups were 3, Group 4 (output-oriented) significantly outperformed Group 2 (input-oriented).

4. Discussion

The first finding of the study was related to the relative effectiveness of input/ output-oriented word-focused tasks with different

involvement loads on learners' vocabulary knowledge of meaning. The fndings of between-group comparisons revealed that the study groups performed differently both on the IP and DP, refecting that the gains were not equal for all the study groups. The results of the study indicated that the participants who completed Task 4 (sentence-making) performed remarkably better that the participants who completed Tasks 1, 2, and 3, on both the IP and DP. In other words, a signifcant advantage was found for Task 4 over the other word-focused tasks. This finding is in accordance with the results obtained by Ellis and He (1999), Fuente (2006), and Webb (2005) on the acquisition of L2 vocabulary and the role of pushed output.

The results of this study indicated that sentence production, as an output- oriented task with an involvement load of 3, contributed to very large gains in learners' vocabulary knowledge. The first plausible explanation for the stronger effect of output-oriented tasks is learners› attention to form. This explanation is very much in line with Toth›s (2006) argument that the learners› internal language processes engaged during the output-oriented tasks might have yielded greater benefts for learning than those of input-oriented tasks. As Izumi (2002) states, output-oriented tasks facilitate L2 development by forcing learners to refect on the L2 forms. The second explanation which might account for the superiority of output-oriented tasks to input-oriented tasks is task demand. Output-oriented tasks involve both processing input and production, while input-oriented tasks involves only input processing. Toth (2006) argues that in output-oriented tasks learners receiving instruction have to respond by encoding their own meaning, but in input-oriented tasks learners have to respond to input by signaling its meaning. The last possible explanation for the superiority of sentence-making task is that this task, as an effective word-focused task, has been used in the

process of learning L1 (Persian) words in elementary schools. Most Iranian EFL learners have had the experience of using sentence-making task in their instructional programs in the process of learning their first language. According to Ellis (2003), familiarity with the task is one of the factors that may promote learning more effectively. Sentence-making provide opportunities for more elaborate processing of the target words in the process of language learning.

The most important finding was related to the signifcance of differences among the study groups as far as the involvement load and input/output orientation of tasks was concerned. The findings of between-group comparisons revealed that the study groups performed differently both on the IP and DP, refecting that the gains were not equal for all the study groups. The study led to interesting findings, some of which concurred while others contradicted the ILH.

The results of the IP revealed that in the receptive test of meaning the participants who completed Task 4 (output-oriented task with an involvement load of 3) outperformed the participants who completed Task 1 (input-oriented task with an involvement load of 2) and Task 2 (input-oriented task with an involvement load of 3). The signifcance of difference between Tasks 1 and 4 partially support the ILH in that both the involvement load and input/output orientation of tasks might have signifcant roles in task effectiveness. However, the signifcant difference between Tasks 2 and 4, and lack of signifcant difference between Tasks 1 and 2 and Tasks 3 and 4 indicated the signifcant role of input/output orientation of tasks.

The results of the DP indicated that in the receptive test of meaning, the participants who completed Task 4 performed remarkably better than those who completed Tasks 1 and 2. The superiority of Task 4 in comparison with Task 1 revealed a signifcant role not only for the involvement load but also for the input/ output

orientation of the tasks. However, the signifcance of difference between Tasks 4 and 2, and lack of signifcant differences between Tasks 1 and 2, and Tasks 3 and 4 revealed a signifcant role for the input/output orientation of tasks as a determining factor in task effectiveness.

The fndings indicated that, in addition to the involvement load of tasks, in- put or output orientation of a task is a determining factor in task effectiveness. If involvement load were the only determining factor in task effectiveness, irrespective of whether the task was input or output-based, the two conditions would yield similar retention results. However, the lack of signifcant difference between Group 1 and 2 and the outperformance of Group 4 in comparison with Group 2 (both with an involvement load of 3) revealed that involvement load was not the only determining factor in task effectiveness. Similar to Folse's (2006) fndings, the results of the study were in contrast with the predictions of the ILH. This revealed that in addition to the involvement load, other factors should be considered in determining task effectiveness.

Unlike previous studies (Fuente, 2006; Hulstijn & Laufer, 2001; Keating, which fully support the hypothesis, the results of this study partially sup- port the ILH. Based on the hypothesis it was anticipated that Task 2 (looking up new words in a dictionary) with an involvement load of 3 to be as effective as Task 4 (sentence-making) with an involvement load of 3, However, the results indicated that it was the other way round, in that Task 4 resulted in higher gains compared to Task 2. One possible explanation for this is that numerical values to the motivational and cognitive elements of the ILH, which in turn yield the in- volvement index, may not carry the same weight or may have been roughly quantifed. In this hypothesis, the amount of the involvement load has been conceived as the sum of the pluses (of need, search, and evaluation). For example, a task consisting of + need, + search, and - evaluation has the same involvement load as a tasks consisting of + need, - search, and + evaluation. Yet all three factors may not be equally important for vocabulary learning. This explanation is very much in line with Kim›s (2008) argument. As Kim states, «it is possible that all three components might not be equal in contributing to vocabulary learning» (p. 313). Along the same line, it might be claimed from the results of this study that each individual component may have a higher or lower variable involvement load. Fur- thermore, the fndings are in accordance with what has already been achieved by Folse (2006). Contra the predictions of the hypothesis, he found that using tasks with strong evaluation was equally as effective as tasks with moderate evaluation. The findings of this study also provided insights for the way that the construct of ILH is related to the Input and Output Hypotheses. The superiority of Task 4 (sentence making) as an output-oriented task may seem to support Swain›s Output Hypothesis (1996), given that the sentence-making task required the learners to stretch their linguistic resources and notice language forms and elaborate on them. The fndings of the study supports Swain's (1993) claim that understanding new forms is not enough and that learners must also be given the chances to use them. Output production induces learners to refect upon language form and this makes acquisition more likely to occur. Swain's (1996) theoretical standpoint is that, without pushed output, learners engage in input comprehension, which does not guarantee future processing of linguistic form that leads to acquisition. One explanation is that output production allows for deeper processing of the L2 words by helping learners to establish more productive meaning-form connections through multiple opportunities for output production of target words. According to Swain (1998), one of the major functions of output is promoting learners' noticing and enabling them to notice a gap in their existing knowledge

brought to their attention by external or internal feedback. The fndings of the study are in line with Gass and Selinker's (2008) argument that input alone is not suffcient for acquisition and output production has an effective role in language learning. The results of the current study indicate that the involvement load may well not function independently of the task type, that is, the input or output orientation of the tasks. In other words, the processing load brought to bear by task type may well affect word retention, a point needing further empirical studies.

Regarding the durability of the word-focused tasks, similar to the findings of Fuente (2006), Keating (2008), and Webb (2005), the results indicated that the improvement for the groups remained signifcant between the pretest and the DP; however, the groups' mean scores showed some decrease on the DP as compared with their mean scores on the IP. One possible explanation that is very much in line with Hulstijn and Laufer›s (2001, p. 274, footnote 20) argument is that "one expects a decline in knowledge over time in the absence of rehearsal or additional exposure to the target words." The participants in this study were not exposed to the target words between testing intervals. Thus it is not surprising that there would be a decline in knowledge for the task that initially showed the greatest gains.

5. Conclusion

Learning a second or a foreign language involves mastering a great number of words. Learning vocabulary through incidental, intentional, and independent approaches requires teachers to plan a wide variety of activities and exercises. The amount of emphasis that teachers and programs decide to place on any given activity will depend on the learner›s level and the educational goals of the teacher and the program.

In investigating the superiority of some word-focused tasks over others and

following the ILH, the main goal of the study was to make a comparison between input and output-oriented tasks with identical involvement loads to see which one plays a more determining role in task effectiveness. It was found that the output-oriented task with high involvement loads produced more effective and more du- rable effects. It can be concluded that output-oriented tasks are more effective in improving the vocabulary knowledge of learners. As Fuente (2006) states, the output production tasks promote attention to form and meaning and help learners make form-meaning connections. He concludes that word-focused tasks can be manipulated so that output production is warranted seem to be sound pedagogical tools for the acquisition of L2 vocabulary items from a task-supported language teaching approach. The most signifcant conclusion drawn from the present study was that, unlike previous studies which rather unambiguously endorsed the ILH, involvement load is not the only determining factor in tasks effectiveness. Rather, in addition to the involvement load of word-focused tasks, the input/output-orien- tation of tasks is also important in determining the effectiveness of a word-focused task. It can be concluded that although Laufer and Hulstijn's (2001) proposal is conceived of as an attempt to stimulate researchers as well as practitioners to operationalize traditional general labels such as noticing, attention, and motivation into task-specifc components, it has long way to go before it achieves its full potential. The findings indicated that there may be more than three determining factors in vocabulary learning. It can be concluded that tasks type, i.e., input or output orientation of a word-focused task, may be also a determining factor. In other words, the involvement index may well not function independently of the task type for vocabulary instruction.

The conclusion that output-oriented tasks were more effective in improving the vocabulary knowledge of learners in this

study adds further support to the widely-acknowledged roles of output and the functions that Swain (1993, 2000) has listed for the output. It seems that involvement load is not a determining factor across all languages and for all learners and, consequently, its beneficial effects should not be over generalized. It is worth mentioning that the greater role of output-oriented tasks in this study does not obviate the role of involvement load in task effective- ness.

The findings of this study might provide a useful tool for language instructors and educators in their selection of effective word-focused tasks. Since vocabulary is considered one of the important components of reading comprehension ability of learners, it might be necessary for teachers to include some word-focused tasks in General English courses where the primary emphasis is placed on the knowledge of the reading comprehension. Furthermore, the fnding that output- oriented tasks with high involvement loads were more effective for vocabulary learning than input-oriented ones with the same involvement load is an encouraging fnding for instructors and researchers who are eager to know in which ways instructional programs might foster the acquisition of target words. Although-input-based tasks seems easier and less time-consuming than output-based tasks, given the superiority of sentence-making as an output-based task with high in-volvement load in this study, it seems plausible to use output-oriented tasks with high involvement loads in General English classes. It does not seem reasonable to completely put aside output-based tasks only because they might be time con-suming or might put too much work on learners' shoulders. The inclusion of such tasks might be crucial to effective vocabulary knowledge.

Similarly, the conclusion that output-oriented tasks with high involvement loads were more effective for vocabulary learning than input-oriented ones with the same involvement load is an encouraging fnding for autonomous learners. In other words, autonomous learners can be informed about the effectiveness of different word-focused tasks so they can make strategic decisions concerning the selection of the most appropriate tasks for improving their vocabulary knowledge. Furthermore, the results of the current study might provide useful insights for the developers of instructional materials and syllabus designers in their selection of effective word-focused tasks in EFL General English materials. It can hardly be denied that an adequate and suffcient vocabulary knowledge lead to good com- prehension. Thus, given the importance of vocabulary in EFL General English classes, any word-focused task that helps learners to develop their vocabulary knowledge would certainly be welcomed.

And finally, this study could possibly lay the groundwork for a great deal of research to touch on the effect of different word-focused tasks on various aspects of vocabulary knowledge. Many other questions may be raised in relation with the fndings of this study. It seems necessary to provide more precise defnitions of the involvement components or add new components to the three proposed by Laufer and Hulstijn (2001), and provide a theoretical underpinning for each of them.

NOTE: The article is based on a thesis submitted in partial fulfllment of the requirements for the degree of Ph.D. in TEFL.

References

Arey, D., Jacobs, L. C., & Razavieh, A. (1996). Introduction to research in education (5th ed.). Florida: Holt, Rinehart & Winston. Inc.

Decarrico, J. S. (2001). Vocabulary learning and teaching (3rd ed.). In M. Celce. Murcia (Ed.), Teaching English as a second or foreign language (pp. 285-299). Boston: Heinle & Heinle.

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

Ellis, R. (2003). Task-based language learning and teaching. Oxford: Oxford University Press.

Ellis, R., & He, X. (1999). The roles of modifed input and output in the incidental acquisition of word meanings. Studies in Second Language Acquisition, 21, 285-301.

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

Fuente, M. J. (2006). Classroom L2 vocabulary acquisitions: Investigating the role of pedagogical tasks and form-focused instruction. Language Teaching Research, 10(3), 263-259.

Garcia, G. E. (1991). Factors infuencing the English reading test performance of Spanish-speaking Hipanic students. Reading Research Quarterly, 16, 371-392. Gass, S. M., & Selinker, L. (2008). Second language acquisition: An introductory course (3rd ed.). NewYork: Routlege: Taylor & Francis Group.

Hazenberg, S., & Hulstijn, J. H. (1996). Defning a minimal receptive second language vocabulary for non-native university students: An empirical investigation. Applied Linguistics, 17(2), 145-163.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & H. Long (Eds.), The handbook of second language acquisition (pp. 349-381). Oxford: Blackwell.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. Language Learning, 51 (3), 539-558.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. Studies in Second Language Acquisition, 24(4), 541-577.

Jafarpur, A. (1992). A course in language testing. Tehran: Payame Noor University Press.

Joe, A. (1995). Text-based tasks and incidental vocabulary learning. Second Language Research, 11, 149-158.

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.

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. Language Learning, 58(2), 285-325. Krashen, S. D. (1985). The input hypothesis: Issues and implications. Oxford: Pergamon Press.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Bejoint (Eds.), Vocabulary and applied linguistics (pp.126- 132). London: Macmillan.

Laufer, B. (1998). The knowledge of passive and active vocabulary in a second language: Same or different? Applied Linguistics, 19 (2), 225-271. Laufer, B. (2005). Focus on form in second language vocabulary learning. EUROSLA Yearbook, 5, 223-250.

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

Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. Language Testing, 16(10), 36-55.

Nagy, W. E., Herman, P., & Anderson, R. C. (1985). Learning words from context. Reading Research Quarterly, 20, 233-253.

Nation, I. S. P. (1993). Vocabulary size, growth, and use. In R. Schreuder & Weltens (Eds.), The bilingual lexicon (pp. 115-134). Philadelphia: John Benjamins.

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

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady T. Huckin (1997), Second language vocabulary acquisition: A rationale for ped- agogy (pp. 174-200). Cambridge: Cambridge University Press.

Schmidt, T. (1990). The role of conciousness in second language learning. Applied Linguistics, 11(2), 129-59.

Swain, M. (1993). The output hypothesis: Just speaking and writing are not enough. The Canadian Modern Language Review, 50, (2), 158-164.

Swain, M. (1996). Three functions of output in second language learning. In Cook & B. Seidlhover (Eds.), For H. G. Widdowson: Principles and practice in the study of language. Oxford: Oxford University Press.

Swain, M. (1998). Focus on form through conscious refection. In C. Doughty J. Williams (Eds.), Focus on form in classroom and second language acquisition (pp. 64-81).

Cambridge: Cambridge University Press. Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), Sociocultural theory and second language learning (pp. 97-114). Oxford: Oxford University Press.

Toth, P. D. (2006). Processing instruction and a role for output in second language acquisition. Language Learning 56(2), 319-385. Webb, S. (2002). Investigating the effects of learning tasks on vocabulary knowledge. Unpublished doctoral dissertation, Victoria University of Wellington.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing of word knowledge. Studies in Second Language Acquisition, 27, 33-52.

Wesche, M., & Paribakht, S. (1998). The infuence of task in reading-based L2 Vocabulary acquisition: Evidence from introspective studies. In K. Haastrup A. Viebrege (Eds.), Perspectives on lexical acquisition in a second language. Lund: Lund University Press.