The Impact of Task-type Based Vocabulary Instruction on Incidental Word Retention

Elaheh Touti

University of Applied Science and Technology (UAST), Tehran, Iran

Roohollah Maleki

Department of English Language, Shahid Sattari Aeronautical University of Science and Technology, Tehran, Iran

Abstract—**Having a good knowledge of vocabulary is a passport to helpful communication. Both theoretical and empirical investigations in the area of L2 incidental vocabulary were taken into account through executing a study comparing the impacts of two types of lexically-oriented tasks at text level (i.e. gap-fill & writing) on the extent of EFL learners' recollection of newly-encountered words. Pursuing this purpose, 64 Iranian EFL learners at intermediate level were selected and then assigned into two experimental groups labelled fill-in group and writing group. This study includes a research question addressing "do lexically-oriented tasks at text level provide better incidental vocabulary recall than gap-fill tasks do? This study, due to the absence of any control group, applied a quasi-experimental design. Statistically speaking, both quantitative and qualitative data were provided. This study used a t-test to compare the mean scores of the two groups. The writing group finally benefited from the outcomes of the study due to the innate mental investment aroused by being engaged in the writing task. The contingency of task engagement rests on the three contributive components (need, search, and evaluation) which could yield the conducive involvement.**

Index Terms—**task-type, incidental words, word-focused tasks, retention rate**

I. INTRODUCTION

Vocabulary occupies the broadest zone of linguistic knowledge in SLA, and thus highly prone to attrition, both in individuals and in language communities (Stringer, D., & Bardovi-Harling, K., 2010, Gass, 2004). Nowadays, the significance of vocabulary manipulation both in the acquisition of one's native language and domineering a foreign language is universally favoured (Morra & Comba, 2009). Vocabulary learning has been taken as an indispensable component in acquiring a second language (L2), since learners should hold vast lexical knowledge to speak, to read, to write, and to gain rich knowledge in L2 (Schmitt, 2008). Hulstijn et al. (2005) believe that unfamiliarity with the meaning of the words appearing in a text seriously paralyze comprehension. Moreover, lexicon constitutes the backbone of language absorption and language utilization (Hunt & Beglar, 2005) In a word, the importance of vocabulary to the English learner is second to none.

Almost every language teaching deals with inviting its learners to grab the knowledge necessary to guarantee productive and communicative use of target language. This purpose could at least seem to be unattainable by only resorting to direct teaching which requires applying the most traditional way of vocabulary instruction. Excessive dependence on direct mode of teaching lexicon, nevertheless, has been impugned due to investing ample time and yielding limited lexical rewards. Direct instruction also, due to a great demand of time and attention to be focused on multiple aspects of language skills, may not solely result in expected uptake and learners' need to such instruction is not assured.

In view of the crucial participatory role of lexicon, it is astonishing that naturally scant class time has been allocated to lexical instruction in general and incidental vocabulary in particular or in case of much time allocation, teaching vocabulary is prone to employing less appropriate useful vocabulary tasks. Vocabulary deficiencies are a primary cause of academic failure, which would impede students' ability to succeed and master a thorough comprehension to overcome communicative obstacles. Furthermore, learners' lexical repertory could be escalated through employing effective tasks inducing sufficient mental efforts conducive to the expected vocabulary retention. The lack of such purposeful tasks in fact prevents students from being involved in mental efforts and making lexical associations that guarantee solid storage and efficient retrieval.

In the realm of vocabulary acquisition, researchers concern with two principles: intentional and incidental absorption of lexical items pursuing direct teaching of vocabulary and dealing with content respectively (Hulstijn, 2001; Nation, 2001). Nonetheless, instructing lexical items through utilizing an intentional module of teaching addresses proceeding with tasks primarily seeking lexical growth. However, learning vocabulary predominantly through extensive reading is labeled incidental learning of vocabulary (Huckin & Coady, 1999; Robinson, 2005; Nakata, 2008). Vocabulary learning tasks contribute greatly to boosting, fortifying, and improving the learners' knowledge of new words.

Two of the most commonly used types of written vocabulary tasks in ESL and EFL classrooms are blank fill-in (i.e. cloze) and writing tasks. Hulstijn and Laufer (2001) and Folse's (2000) studies aligned with limited number of investigations embarked on examining the two tasks of gap-fill and writing. The tasks they tried to examine included a) dealing with comprehension inquiries, b) gap-fill and c) writing. The outcomes verified the superiority of writing, due to inducing higher mental efforts as its inborn feature, over the other two tasks in promoting learning and retention of incidental words. In fact, Hustijn and Laufer compared the two types of tasks at the text level, while Folse conducted his study aiming at comparing the two types of activities at the sentence level. Different outcomes were attained from both studies with regard to the effectiveness of the two types of treatment (i.e. fill-in-the-blank and writing) (San Mateo Valdehita, 2003; 2004, p.32). Laufer (2001) indicated that better results could be obtained if reading were accompanied by word-focused tasks. Rott, Williams and Cameron (2002) asserted that embedding vocabulary items in tasks using dictionary or getting benefit from glossed provisions of the words could culminate in better lexical absorption than those presented in reading-alone exercises. Therefore, language learners and language teaching professionals are looking for purpose-built tasks to widen learners' lexical repository efficiently.

The focus of this study is captured on discovering the effects the two types of lexically-oriented tasks at text level (i.e. gap-fill and writing) have on the EFL learners' recall of unknown words and consequently looking for the most influential task in terms of representing better memory traces in the learners' recall of newly-encountered lexical items. Moreover, the findings of this study targets a specific type of audience, namely teachers, textbook writers, materials designers, and curriculum leaders who should attempt to take the advantage of the results addressing the goal of any language teaching program, in improving their lessons and syllabuses in order to empower learners acquire new words more quickly and efficiently. The study, also contributes increasingly to a comprehensive writing theory which puts lots of emphasis on the miraculous role of the task of writing in enhancing the retention of the incidental vocabulary.

With respect to the effect of employing vocabulary-based tasks on learners' retention rate of unknown words, previous studies (Hulstijn & Laufer, 2001) conducted a study aiming at investigating the effect of three tasks of reading comprehension with marginal glosses, reading-based gap-fill in questions, and composition writing on learners' recall of unknown words. However, no research has so far been conducted comparing specifically the effect of the two tasks of fill-in and writing at text-level on learners' retention rate of incidental words. The need for conducting such a research in literature in fact captures the focus of this study to address this concern meticulously in the hope for to coming up with findings conducive to second language instruction and learners' progress in incidental vocabulary recall.

II. THE INVOLVEMENT LOAD HYPOTHESIS

Schmidt (2008) maintained that the more immense learners engross in novel lexically-oriented tasks, the higher their chance will be in word absorption. In their search for an operationalizable definition of the depth of processing theory (Craik & Lockhart, 1972), Laufer and Hulstijn introduced the Involvement Load Hypothesis consisting of three principal constituents: need, search and evaluation. The first constituent stimulates motivation addressing a mentally-free facet of involvement. There exist two levels od prominence for need, including moderate and strong distinguished regarding external-internal demarcation. For instance, the moderate level of need arises by an external trigger (e.g. the need to diagnose the meaning of a lexical item in a sentence as has been asked by the teacher). The strong level of need arises as learners proceed with a self-oriented attempt in dealing with a task. (e.g. the need to check the word from a dictionary while composing a piece of writing).

Two levels of prominence belong to the evaluation constituent. The employment of a novel lexical item within an intended sentence is required for the moderate level to be aroused. However, generating authentic sentences addresses the strong level of evaluation, since it allows learners to recognize the way in which words can be assembled coherently. Hulstijn and Laufer (2001) asserted that all of the three constituents may not gather together concurrently while dealing with a reading activity, and what includes the involvement load supports the synthetic presence of these constituents with their levels of prominence. Involvement Load Hypothesis holds that learners display various responses through being engaged in different tasks. Nevertheless, comparing various tasks and deciding on their level of involvement unrealistically is beyond our ability. Consequently, researchers tried to shift the theoretical concept of involvement load into a quantifiable notion of task-induced engagement. Hulstijn and Laufer (2001) declared that the primary tenet concerning involvement load entails the contingency of the recollection of unfamiliar words on the level of task engagement in deciphering these words. In their attempt to operationalize depth of processing, Hustijn and Laufer proposed an index for different levels of task engagement, in which 0 belongs to the nonattendance of a constituent, 1 belongs to a moderate attendance of a constituent and 2 belongs to a strong attendance of a constituent. The concept of involvement can be measured empirically by designing tasks accompanied by different levels of need, search and evaluation seeking to calibrate learners' extent of recall of the novel words. For instance, to witness any relationship between the degree of task engagement and word recall, tasks enjoying various indexes of involvement can be undertaken by learners and the outcomes can be ultimately examined. Below discusses the previous empirical studies conducted on involvement load hypothesis.

III. EMPIRICAL STUDIES ON INVOLVEMENT LOAD HYPOTHESIS

The emergence of the concept of Involvement Load Hypothesis encouraged Hulstijn and Laufer (2001) to conduct a study pursuing the aim of investigating the impact of the degree of task engagement on immediate and delayed recall of 10 unfamiliar lexical items by advanced EFL learners in two different experiments. Three tasks accompanied by diverse levels of involvement were compared in their study. The employed tasks were a reading comprehension equipped with marginal descriptions, reading comprehension aligned with gap-fill, and composing a piece of writing utilizing the candidate words to which 1, 2, and 3 indexes of task engagement were allocated respectively.

Regarding the degree of task engagement, the first task induced moderate need. However, it came up with the nonattendance of search or evaluation constituent. Therefore, the index of task engagement equaled 1 (1+0+0). The second task came with the moderate attendance of need constituent and an absence of search constituent and the presence of moderate evaluation, since the lexical words were expected to be inserted in a set context. Thus, the index of task engagement belonged to this task equaled 2 (1+0+1). Students in the third task were asked to compose a piece of writing employing the ten target lexical items the meanings of which were provided marginally aligned with example sentences. The third task came up with moderate need , an absence of search and the presence of strong evaluation, since students made attempts to embed the target words in their piece of writing, therefore, the index of task engagement equaled 3 (1+0+2). Hulstijn and Laufer (2001) with respect to the involvement load hypothesis documented that task 3 (i.e. lexically-accompanied composition writing) was reported to trigger the highest degree of recall of the candidate words. The other two tasks (i.e. reading comprehension aligned with gap-fill and reading comprehension) were reported to yield less and least degree of recall respectively.

Productive use of vocabulary (e.g., through writing sentences), as proposed by Hulstijn and Laufer endorses the high involvement load induced and thus facilitates lexical learning. In this sense, the results are also in harmony with previous literature in concluding that deeper involvement with vocabulary-based tasks contributes measurably to the lexical acquisition (Hill & Laufer, 2003; San Mateo Valdehita, 2003, 2004; Laufer, 2004; Rosa & Leow, 2004; Sanchez, 2004).

Kim (2008) conducted an experiment with 40 young adult English as L2 learners, from nine countries, who spoke 18 different L1s. The purpose of the study addressed the investigation of whether the effect of the two activities (e.g. writing original sentences or a composition employing the target words in both cases) enjoying equal theoretical mental effort or involvement load (Hulsijn & Laufer, 2001), is the same on the concurrent association of novel words with the mental lexicon and also on their recall. Kim (2008) taking the involvement load hypothesis into account believed that the same degree of engagement in two lexical-oriented tasks should engender identical outcomes of lexical absorption. An adaptation of the Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1993) was used in the post-tests to gauge vocabulary knowledge; the first one was administered immediately after the learning task and the delayed post-test two weeks later. The analysis of results revealed the difference between the two tasks (i.e. composition and sentence writing), involving "a moderate *need*, no *search*, and strong *evaluation*", (Kim, 2008, p. 310) was not significant; thus these activities both yielded equal effect in enhancing vocabulary in L2.

By stating the deficiencies of Hulstijn and Laufer's (2001) study, Keating (2008) in a study examined the impact of higher mentally-induced tasks on less competent learners' same lexical absorption on passive and active tests. Three tasks with various degrees of involvement were the target of Keating's study with one of which Spanish learners were required to deal with. The treatment tasks consisted of reading comprehension equipped with marginal descriptions, reading comprehension aligned with gap-fill, and sentence generation accompanied by the candidate words. Meeting the purpose of Keating's study, the participants underwent the third task due to possessing higher mental efforts. Having completed the task, learners were administered two passive and active tests after two weeks and their recall of the candidate words was gauged accordingly. The outcomes obtained from both immediate and delayed passive tests though not absolutely supported the involvement load hypothesis demonstrated the most immense recall and the highest score in terms of reading comprehension accompanied by gap-fill and sentence generation accompanied by the target words learner were expected to employ, compared to reading comprehension with marginal descriptions. Nonetheless, the task of sentence generation accompanied by candidate words did not exhibit higher efficacy than reading comprehension accompanied by gap-fill. Contrarily, the results reported elevating lexical recall for the task of sentence generation equipped with candidate words and the task of reading comprehension plus gap-fill compared to the task of reading comprehension with marginal descriptions. Nevertheless, learners in the task of sentence generation accompanied by the candidate words did not display considerable progress than the participants dealing with the task of reading comprehension with marginal descriptions and the task of reading comprehension accompanied by gap-fill.

In a nut shell, the findings of the study performed by Keating (2008) manifested that the involvement load hypothesis may be applied universally to target less competent learners and may likewise leave an impact on the learners' passive and active word recall. Following Keating (2008), Kim (2011) examined the Involvement Load Hypothesis in an ESL setting, across various exercise types and proficiency levels with a controlled time on task. In his first experiment, he investigated the efficacy of three tasks with different involvement loads within two different proficiency levels. In each proficiency level, learners randomly completed one of three tasks: reading, gap-fill and composition. In order to assess L2 learners' initial learning and recall of target words, two immediate and delayed post-tests were executed. The results of both post-tests manifested the highest involvement index of 3 for the Composition group, the index of 2 for the Fill-

in group and the index of 1 for the Reading group. Nevertheless, the Fill-in participants on the delayed post-test obtained considerably elevating scores compared to the Reading participants.

Concisely, Involvement Load Hypothesis was reported to be completely practical in terms of the post-test administered after quite a short time and partially practically in terms of the post-test administered later. In his second experiment, Kim (2011) examined whether two tasks enjoying the same involvement loads affected similarly learning of target words. The author, therefore, compared the writing composition (index = 3) with the writing sentence task (index = 3). The results of both post-tests displayed that these two tasks with equal involvement loads affected similarly the initial learning and retention of target words across two different proficiency levels, a claim which was supported by the Involvement Load Hypothesis.

The impact of written tasks on second language word recall was examined by Folse (2006). 154 ESL participants comprised the sampling population of his study. They were selected from four universities in the U.S. They underwent 18 unfamiliar lexical items in three diverse conditions. The tasks employed in Folse's study consisted of one gap-fill task, three gap-fill tasks, and one sentence generation task. The two former tasks were recognition-based and the latter was production-based. The unrehearsed post-test was administered to gauge the extent of recall in the various circumstances. Through running a repeated ANOVA measurement, it was manifest that the participants undertaking the three tasks of gap-fill gained the highest score and consequently their word recall encountered an escalation. The significance of dealing with excessive retrievals of target lexical items in a task is worthy of attention in the zone of L2 lexical absorption.

IV. INCIDENTAL VOCABULARY LEARNING AND TEXT-BASED WORD-FOCUSED TASKS

Many scholars (Fraser, 1999; Huckin & Coady, 1999) argued that a sizeable amount of vocabulary learning finds its birth in reading tasks. It has been manifested that incidental vocabulary learning through employing reading tasks only brings about an inconsiderable rate of retention. This notion is supported by some experts such as, Pressley, Levin & McDaniel (1987), who claimed that learners are able to capture vocabulary meanings by the help of context, but this mainly doesn't lead to a substantial amount of retention of meanings. Due to the insufficient vocabulary gain through an absolute reliance on reading, a need for effective tasks to fortify the students' existing knowledge is seriously urgent. This inadequacy emanated from the sole reliance on reading-only tasks is evident in empirical studies (Paribakht & Wesche, 1999; Zimmerman, 1997), displaying the fact that L2 vocabulary recall escalates in case of being engaged in written vocabulary tasks. As Stoller and Grabe (1993) declared, the effects of incidental learning can be increased by means of employing reading plus word-focused tasks.

A number of studies has demonstrated that remarkable vocabulary gains were expected from subjects' dealing with cognitively demanding tasks that render deeper levels of processing and greater mental effort of encoding the vocabulary items necessary (Hulstijn & Laufer, 2001; Webb, 2000; Mateo Valdehita, 2004). This provokes a principal inquiry regarding characteristics of the most influential written vocabulary tasks which are described below.

V. METHODOLOGY

A. Participants

The participants who were selected from different language institutes in Tehran were composed of 90 young intermediate EFL learners (females) among whom, due to meeting homogeneity, 64 were targeted. The participants were from sixteen to thirty years of age. Then, they were randomly assigned to two 32 homogeneous experimental groups. The participants' mother tongue was Persian.

B. Materials

The whole participants selected for the study (i.e. 90 participants) received the 2010 version of PET, with its speaking part eliminated due to time constraint. To meet homogeneous language proficiency levels, the whole participants underwent a preliminary English test (i.e. PET). The reliability of PET was computed as .90 (see discussion section for detailed statistical analysis). From the whole population, 64 participants were recognized as being homogenous and consequently selected as the main participants of the study. The involved participants were then randomly fallen into two 32 experimental groups. Next, they were administered the piloted pre-test. After that, they underwent the treatment, and the piloted post-test.

1. Teacher-made Diagnostic Vocabulary Test as the Pre-test

The pre-test was composed of eighty two multiple choice teacher-made incidental (i.e. unknown) vocabulary items on *Transportation and City Services*. It was, prior to be administered to the main participants, piloted with another group of participants enjoying the same characteristics (i.e. the same age and proficiency level) as the main participants of the study. The candidate words were accompanied by a small number of familiar words to mitigate the students' frustration and consequently heighten their confidence. In this phase, each target word was embedded in a multiple choice sentence so that for each item the participants were to mark the best option in their answer sheet.

During the piloting procedure, twelve vocabulary items, enjoying the highest number of correct answers, were excluded from the original list of the target words (i.e. eighty two words). The modified list of the target words

consisting of seventy multiple choice vocabulary items were then administered to the main participants of the study, comprising the two thirty two homogeneous experimental groups, to recognize their least known words. Having checked the participants' pre-test sheets, the teacher-researcher recognized twenty eight vocabulary items as their unknown lexical items. The final list of incidental words (i.e. the participants' diagnosed least known words consisting of twenty eight words), was considered to be employed as the post-test in the multiple choice format. The reliability of the test was then estimated (see discussion section for statistical analysis). Each paper sheet was equipped with a PIN number asked to be remembered by the participants. This in fact made the participants' post-test record tracking easier for the teacher-researcher.

2. Teacher-made Achievement Test as the Post-test

The post-test of the study consisted of twenty eight unknown words which were recognized as the participants' least known words in the pre-test. The post-test, was piloted with another group of participants enjoying the same characteristics (i.e. the same age and proficiency level) as the main participants of the study. In fact, it was piloted for the reliability estimation prior to be administered to the main participants. The intended words employed in the fill-in and writing tasks were utilized as the post-test in the multiple choice format. These twenty eight words were divided into seven words that were distributed across the four tasks assigned for each experimental group in four treatment sessions (i.e. together eight treatment sessions). The post-test designed in multiple choice format was derived from the same seven words utilized in each assigned task practiced by the participants of the two experimental groups in the treatment sessions. It was administered to each group immediately after the task completion and was, following the same format as that of the pre-test, aligned with an answer sheet.

VI. PROCEDURE

The study was conducted in three different language institutes in Tehran: kiyan, Farzanegan, and Safiran in the fall semester in 2011. This study consisted of administering the 2010 version of PET to the two experimental groups (i.e. Fill-in and Writing) to confirm homogeneity. The reliability of PET was computed as .90 (see discussion section for detailed statistical analysis). The speaking section of the language proficiency test due to the time constraint was excluded. The rest of the study comprised eighty two multiple choice teacher-made diagnostic vocabulary items, that seventy of which after being piloted were adopted to be administered as the pre-test with its reliability computed by Cronbach's Alpha method as .87. Also, the post-test included twenty eight vocabulary items which were selected and recognized as the participants' least known words in the pre-test. Each post-test comprised of seven of the target words designed in the form of multiple-choice format which was administered separately and immediately after task completion in each treatment session. The reliability of the post-test as calculated by Cronbach's Alpha method was .80. The target incidental vocabulary items were on the topic of *Transportation and City Services* and included non-textbook-bound intermediate lexical items adopted from a variety of sources, such as Longman Dictionary of Contemporary English and the website www.enchantedlearning.com, as well as several sources, including Kelly Wingate Publications on comprehension skills and Wikipedia, from which the treatment task sheets were selected. Sixty four participants, adhering to the principle of standard deviation: one SD above and below the obtained mean, took part in the study. They were selected and identified as homogeneous out of a large population (i.e. ninety participants) after taking PET. The assumed participants were then equally divided into two thirty two experimental groups of the Fill-in and Writing. Respecting the research moral principles, the participants of the study were kept informed of the aim of the study and invited to acknowledge their consent in advance. The treatment sessions conducted for both tasks were teacher-researcher-oriented.

Regarding the treatment to which eight two-hour sessions (i.e. four two-hour sessions for each group) were allocated, seven of the participants' twenty eight vocabulary items specified as their least known words in the pre-test and also considered to be utilized in the post-test, were employed in the form of four thematically different reading passages, adopted from Kelly Wingate Publications on Comprehension Skills, Wikipedia and some teacher-made reading passages. The passages were accompanied by a word study sheet, including the English definitions of the target words utilized in the tasks with two extra lexical items for the FT group to lower the chance of guestimate and the same seven words for the WT group. The topics and contents of the reading passages were the same for both groups. Also, the target words utilized in the treatment tasks for the two groups were the same. The only difference was that the blanks in the FT task were filled in with the same seven candidate words utilized in the reading-based WT task. Furthermore, learners' attention was grasped by the case words in both tasks. The extraction of the intended words in the FT task and their being bold-printed in the reading-based WT task in fact arose the learners' attention. The two experimental groups underwent completion-type and reading-accompanied writing tasks respectively. Each group experienced a four-week treatment session (i.e. altogether eight sessions) based on the so-called task-type based instruction prior to the post-test.

Each session, immediately after the completion of the tasks, the task sheets along with their answer sheets were collected and the two treatment groups underwent altogether eight post-tests each consisting of seven vocabulary items, derived from the same seven words used in each treatment tasks practiced by the two groups, in multiple choice format that was each administered separately in each treatment session. The participants were assumed to practice the tasks on their own and the teacher-researcher explained the process of dealing with the tasks prior to practicing them. To meet the incidental nature of the vocabulary learning, all treatment sessions, including sessions in which the participants

would undergo practicing the tasks and post-task vocabulary recall tests (i.e. post-tests) administered separately each session immediately after tasks completion to measure the participants' retention of the target incidental words, were held unexpected and without the participants' advance notice.

As for the FT (i.e. fill-in at the text level) treatment group, the participants were expected to read a reading passage from which some words were extracted. The participants were to find out the sought omitted words and settle them in the target blank spaces in the text. Having tackled both the task and the accompanied test administered immediately after the task completion, the teacher-researcher in case of any coming ambiguities regarding the meaning of the words or any other questions asked by the participants, provided lucid explanations of the sought concerns through mutual negotiations. The teacher-researcher in encountering the participants' willingness towards further elaborations encouraged the participants to manifest variety in their output. The rest of the explications were dedicated to practicing the parts of speech and the loud repetition of the target words.

In the WT (i.e. writing at the text level) treatment group to which four sessions were allocated, the participants were, making use of the same seven target words utilized in the FT task, assigned the similar reading passages as those presented to the FT treatment group. Each reading passage was attached with a writing sample sheet on which the participants were expected to write their compositions after they grasped sufficient comprehension of the reading passage. The reading passages were followed by word-focused comprehension questions addressing the target words in the text. After reading the passage, the participants were asked to go through a writing task in which they were presumed to compose a piece of text using the same seven target words which were bolded in the reading passages practiced in the treatment sessions. The participants were expected to create the writing topic themselves. Having finished this step, the participants were then administered a post-test similar as the one administered to the FT treatment group. The test consisted of the same seven words utilized in the WT task following the multiple choice format plus an answer sheet which was administered separately after each task completion in each treatment session. After the completion of both steps (i.e. task and test completion), the teacher-researcher, via peer negotiations, removed any meaning-based obstacles regarding the respective words and all the related concerns through clarification.

Entering the realm of the post-test, the participants of the both experimental groups were administered a piloted post-test each session after the task completion. Generally speaking, the post-test consisted of twenty eight vocabulary items in multiple choice format that the participants went through seven of which in each treatment session assigned for each group. The reason behind separate and immediate administration of multiple post-task tests (i.e. post-tests), after the task completion in eight separate treatment sessions, supported the view that the more the participants were exposed to practicing their own task sheet, the more accurate results could be obtained from their performance.

VII. DATA ANALYSIS AND RESULTS

The research question of this study aimed at finding out whether lexically-oriented tasks at text level provide better incidental vocabulary recall than gap-fill tasks do. In order to answer this research question, independent sample *t*-test was performed.

A group of 90 intermediate EFL learners took PET to select homogeneous participants. The descriptive statistics for PET, as set forth in Table 1, reflects that the mean, median and mode of the PET scores are 44.32, 45.50, and 48 respectively. These central parameters are close to one another other implying that the scores are normally distributed around the mean. Based on the results of PET (Table 1), those 64 students whose scores were one standard deviation (12.64) plus and minus the mean of 44.32 were selected as homogeneous intermediate participants for the main research.

TABLE 1.
DESCRIPTIVE STATISTICS FOR PET

N	Range	Min	Max	Mean	Median	Mode	SD
94	54	14	68	44.32	45.50	48	12.641

We used parametric test since four assumptions (i.e., interval data, independence of subjects, normality and homogeneity of variances) were met (Field, 2009). In fact one-sample Kolmogorov-Smirnov test (Table 2) showed that all sets of vocabulary scores for the fill-in and writing groups on both pre-test and post-test have normal distribution (Sig. > .05).

TABLE 2.
ONE-SAMPLE KOLMOGOROV-SMIRNOV TEST OF NORMALITY

Test	Group	*N*	Mean	Kolmogorov-Smirnov Z	Asymp. Sig. (2-tailed)
Pre-test	FT	32	18.31	1.139	.167
	WT	32	17.41	1.272	.102
Post-test	FT	32	19.53	.567	.905
	WT	32	23.50	.690	.728

Before discussing the results of *t*-test, the descriptive statistics for the fill-in and writing groups were computed (Table 3). A quick look at the table reveals that the mean and standard deviation for the fill-in ($\bar{x}$ = 18.31, *SD* = 8.37) and writing ($\bar{x}$ = 17.41, *SD* = 7.20) groups do not differ highly on pre-test of incidental vocabulary recall.

TABLE 3.
DESCRIPTIVE STATISTICS FOR THE TWO GROUPS ON THE PRE-TEST

Group	N	Range	Minimum	Maximum	Mean	SD
FT	32	37	8	45	18.31	8.372
WT	32	34	8	42	17.41	7.202

Table 4 contains the result of independent *t*-test that was applied to compare the fill-in and writing groups' vocabulary scores on the pre-test. The table shows that the significance level for Levene's Test (.58) is greater than the selected significance level (.05), so the assumption of equal of variances is not violated. Also, Table 4 indicates that *t* value and significance level (*t* (62) = .46, *p* = .64, *p* > .05) are indicative of no significant difference in vocabulary scores for fill-in ($\bar{x}$ = 18.31) and writing ($\bar{x}$ = 17.41) groups, in which the *t*-observed is below than the *t*-critical of 2.00; hence we conclude that the students in the two groups have the same vocabulary knowledge at the beginning of the study.

TABLE 4.
INDEPENDENT SAMPLES TEST FOR VOCABULARY SCORES ON THE PRE-TEST

	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	*t*	*df*	*Sig.* (2-tailed)	Mean Diff.	Std. Error Difference	Lower	Upper
Equal variances assumed	.303	.584	.464	62	.644	.906	1.952	-2.996	4.809
Equal variances not assumed			.464	60.645	.644	.906	1.952	-2.998	4.810

Before presenting the results of *t*-test for the post-test, the related descriptive statistics are laid out in Table 5. According to Table 5, the students in the writing group ($\bar{x}$ = 23.50, *SD* = 3.34) outperformed those in the fill-in group ($\bar{x}$ = 19.53, *SD* = 3.76) on post-test of incidental vocabulary.

TABLE 5.
DESCRIPTIVE STATISTICS FOR THE TWO GROUPS ON THE POST-

Group	N	Range	Minimum	Maximum	Mean	SD
FT	32	15	11	26	19.53	3.767
WT	32	13	15	28	23.50	3.341

Further, another independent *t*-test was run to compare fill-in and writing groups' incidental vocabulary scores on the post-test (Table 6). As can be seen in Table 6, the assumption of equal of variances is met (*Sig.* > .05).

TABLE 6.
INDEPENDENT SAMPLES TEST FOR VOCABULARY SCORES ON THE POST-TEST

	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	*Sig.*	*t*	*df*	*Sig.* (2-tailed)	Mean Diff.	Std. Error Difference	Lower	Upper
Equal variances assumed	.35	.55	-4.45	62	.000*	-3.96	.89	-5.74	-2.18
Equal variances not assumed			-4.45	61.12	.000*	-3.96	.89	-5.74	-2.18

p < .05 = It shows significant difference

In addition, Table 6 reflects that independent *t*-test detected a statistically significant difference (*t* (62) = 4.45, *p* = .000, *p* < .05) in incidental vocabulary recall scores for fill-in ($\bar{x}$ = 19.53) and writing ($\bar{x}$ = 23.50) groups, in which the *t*-observed is above than the *t*-critical of 2.00. Consequently, we strongly reject the null hypothesis of the current study that states, "Lexically-oriented tasks at text level do not provide better incidental vocabulary recall than gap-fill tasks do" and therefore claim that lexically-oriented tasks at text level provide better incidental vocabulary recall than gap-fill tasks do.

We made a bar graph (Figure 1) to graphically show the results on both pre-test and post-test. As the figure illustrates, the students in the fill-in group have performed significantly better than those in the writing group regarding incidental vocabulary recall in the condition that they had the same vocabulary knowledge on the pre-test.
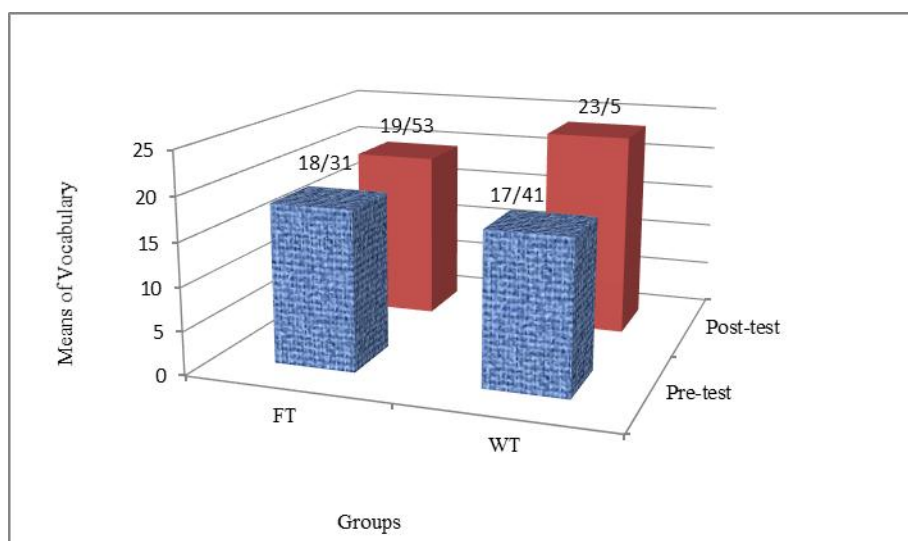

Figure 1. Two Groups' Means on Vocabulary Pre-test and Post-test

The findings of this study strongly support those of other studies conducted by Hulstijn and Laufer (2001), Webb (2005), and Mateo Valdehita (2004) sharing the common view of valuing the benefit of writing tasks and admiring the role of writing at the text level (i.e. WT). Moreover, the findings of this study endorse those of the previous research conducted by Hulstijn and Laufer (2001) in that text-level tasks (i.e. the task of writing) compared to sentence-level tasks (i.e. the task of fill-in) due to inducing higher involvement load, leading to deeper engagement in tasks, and requiring athletic mental efforts to be invested by learners resulted in more robust incidental vocabulary retention.

The study also magnifies the validity of Hulstijn and Laufer's Involvement Load Hypothesis along with its three components of task induced involvement (i.e. need, search, and evaluation) which represents the practicality of general cognitive notions of depth of processing and elaboration in L2 vocabulary tasks. The findings of the study favor Laufer and Hulstijn's claim (2001) that any specific task type— be it input or output—doesn't yield escalation or growth, and the primary feature of a quality task addresses the degree of task engagement.

The chief goal sought by this study dealt with introducing the most purposeful task in terms of learners' lexical absorption. The findings of the study were in line with those of Hulstijn and Laufer (2001), Folse (2006), Keating (2008), and Kim (2011) in supporting the efficacy of the involvement load hypothesis with respect to employing highly cognitive-demanding tasks enjoying higher amount of involvement load and absorbing enormous mental efforts. The verified effect of the involvement load hypothesis, through recruiting tasks enjoying higher involvement load, on the learners' escalation of incidental word gains was transparent in the findings of this study that endorsed the superiority of WT group due to yielding higher scores in incidental vocabulary retention.

VIII. CONCLUSION AND PEDAGOGICAL IMPLICATIONS

The focus of the study is captured upon validating the Hulstijn and Laufer's Involvement Load Hypothesis with its three constituents of need, search and evaluation displaying the degree of task engagement which manifests the operationalization of depth of processing and elaboration circumscribed by L2 vocabulary tasks. Tasks that take various language facets into consideration deserve close attention. Advocated by Hulstijn and Laufer's Involvement Load Hypothesis and demonstrated statistically, tasks with higher involvement load, namely tasks requiring writing, led to a sure-fire breakthrough in this experiment. Furthermore, the incidental acquisition of the target word in a task yielding greater involvement load culminates to a more solid retention of those words.

Besides, the contribution of certain word-focused tasks to vocabulary knowledge and the successive escalation in the retention of that knowledge compassed the focus of this study. In this regard, the context as well as the attention and depth of input retrieval induced by word-focused tasks contribute interactively. This accords with what Laufer and Hill (2000) believed to guarantee a productive teaching highlighting the significance of learners' attention to the instructionally-oriented candidate words while being engrossed in tasks arising lexical enhancement. Comparing the mental processes congenital to the two tasks of writing and fill-in, the conclusion displayed that the former significantly induced deeper processing of the target words than just assembling the patchwork of ideas in an existing context in the latter task (Hulstijn & Laufer, 2001).

To sum up, word-focused tasks were undoubtedly corroborated to result in a substantial vocabulary retention rate overall in assisting second language learners to master an enduring reservoir of newly-met words. The findings of this

study verified those of other relevant (Newton, 1995; Joe, 1995, Pribakht & Wesche, 1999; Zimmerman, 1997) studies which were conducted in the area of vocabulary learning. These tasks basically draw learners' attention to specific vocabulary items to become involved in reflective planning, catenizing thoughts (i.e. chaining words together), and thus creating plausible versatilities to self-generate sequenced discourse.

Pedagogically speaking, the findings of the study espouse employing written word-focused tasks inducing higher mental involvement conducive to better retention of incidental words (Hulstijn & Laufer, 2001). Concerning the obtained results, embedding new words in meaningful contexts represented in a piece of writing in fact nourish learners' mentality in a way to improve their efficacies to utilize words in the future. Moreover, writing activities give the instructional mentor (i.e. the teacher) the opportunity to monitor the students' progress and remove any linguistic obstacles, including lexical hurdles they might confront. Furthermore, the Involvement Load Hypothesis allows teachers and task designers to manipulate task features and predict what task will best deserve academic and curriculum manipulation in terms of fostering incidental vocabulary retention. However, both theoretical and empirical research in the realm of L2 vocabulary learning should be manipulated in an attempt to explore more the nature of the construct of task-induced involvement, with three motivational and cognitive dimensions: need, search and evaluation.

APPENDIX A

Preliminary English Test (PET)

Test 3 Reading Part 1

Test 3

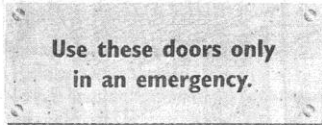## Paper 1 Reading and Writing (1 hour 30 minutes)

Reading
Part 1

Tip
Don't choose the first option you think is right. Look carefully at the meaning of each option before you decide.

Questions 1–5

Look at the text in each question.
What does it say?
Mark the correct letter A, B or C on your answer sheet.

Example:

0   Use these doors only in an emergency.
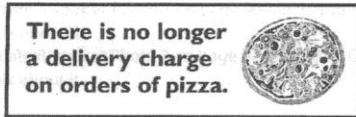
A   You must never leave by these doors.
B   The doors can be used if necessary.
C   Only some people can use these doors.

Answer: 0 | A B C

1   There is no longer a delivery charge on orders of pizza.

A   This restaurant gives a free pizza with every order.
B   You don't have to pay for a pizza delivery.
C   Some orders of pizza cannot be delivered.

2   Passengers without passes should have the correct fare. Drivers are unable to give change.
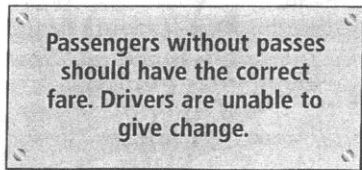
A   Passengers must have a pass or the exact money.
B   Some fares on this bus route have changed.
C   You can only travel on this bus if you have a pass.

40

APPENDIX B

Pre-test
Direction:
Select the best option and mark it in your answer sheet. Write your PIN number in the circle located at the right hand corner of your answer sheet.

1. The Plane was carrying a full…………………..of fuel.
   a) Rate   b) Scale   c) Load   d) Balance
2. The city of Boerne was…………………… by German settlers in the 1840s.
   a) Established   b) Shifted   c) Excavated   d) Removed
3. Due to a strong storm, the ship was…………………… and most of the crews were drowned.
   a) Repaired   b) Shacked   c) Floated   d) Keeled over
4. There wasn't enough water to…………………… the ship.
   a) Move   b) Float   c) Push   d) Draw
5. Last week a tragic accident took place in San Diego………………….. .
   a) Freeway   b) Subway   c) Flyover   d) Escalator
6. He braked suddenly, causing the front…………………… skid on the ice.
   a) Taillights   b) Hubcaps   c) Headlights   d) Wheels
7. The survivors of the ……………………. were found miles away from the Island.
   a) Accident   b) Earthquake   c) Shipwreck   d) Drought
8. Emergency food suppliers were brought in by………………………. .
   a) Lorry   b) Camper   c) Tram   d) Submarine
9. Steward was injured in a…………………………. with another player.
   a) Collision   b) match   c) Kick   d) Game
10. The ship……….................... to the bottom of the sea.
    a) Travelled   b) Sank   c) Reached   d) Fell
11. The bridge is reinforced with huge………………………… pillars.
    a) Steel   b) Iron   c) Wooden   d) Metal
12. He…………………………. the boat into the harbor.
    a) Guided   b) Steered   c) Helped   d) Rode
13. The government decides to……………………………. a new road bridge across the river.
    a) Damage   b) Repair   c) Construct   d) Change
14. A: What are the children doing?
    B: They are having a……………………….. down the river.
    a) Raft   b) Kayak   c) Canoe   d) Punt

APPENDIX C

Post-test
Direction:
Select the best option and mark it in your answer sheet. Write your PIN number in the circle located at the right hand corner of your answer sheet.

1. The sailors use…………………. to control the direction of a boat.
   a) Flag   b) Receiver   c) Rudder   d) Transmitter
2. The……………….. parking on both sides of the streets make the appearance of the city ugly.
   a) Clunkers   b) Architectures   c) Garbage   d) Ruins
3. London's roads and city streets are too…………………. to drive smoothly.
   a) Congested   b) Deserted   c) Polluted   d) Dirty
4. A: Who do you think would win in the…………………. .
   B: I guess Martin would be placed first. What about you?
   A: I don't think so. His boat is old and needs repairs.
   a) Archery   b) Bowling   c) Regatta   d) Fencing
5. The ship's captain didn't manage to steer it forward, because of holes in its fuselage which should be……………… .
   a) Decreased   b) Filled   c) Disappeared   d) Caulked
6. The company decides to…………………. all its latest computers due to a CPU problem.
   a) Recall   b) Brand   c) Equip   d) Produce
7. A boat is…………………. by a small motor.
   a) Controlled   b) Outrun   c) Tracked   d) Propelled
8. In the past people used to travel by…………………. carts, pulled by horses.
   a) Diesel   b) Buggy   c) Locomotive   d) Stroller

9. The vessel was attacked by warship and……………………. .
   a) Returned   b) Keeled   c) Bounced   d) Threw
10. The………………….. of fish stocks is expected to happen in an increase in the amount of poisonous substances.
   a) Preparation   b) Gathering   c) Extraction   d) Depletion
11. In the ship's workshop, ways of designing…………………… boats and ships which are use for hunting and fighting are going to be discussed.
   a) Fowling   b) Cruise   c) Cargo   d) House
12. This traffic jam annoys me. It's best to travel by……………………. .
   a) Spaceship   b) Tour   c) Carriage   d) Tube
13. The school bus was involved in a……………………. with a truck.

REFERENCES

[1]   Craik, F.I.M., & Lockhart, K.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.
[2]   Field, A. (2009). Discovering statistics using SPSS. 3rd ed. SAGE: London.
[3]   Folse, K. (2000). The effect of type of written practice activity on second language vocabulary retention. Unpublished doctoral dissertation, University of South Florida.
[4]   Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, *40*(2), 273-293.
[5]   Fraser, C. A. (1999). LPS use and vocabulary learning through reading. *Studies in Second Language Acquisition, 21*, 225-241.
[6]   Gass, S., (2004). Context and SLA. In B. VanParren, J. Williams, S. Rott and M. Overstreet (Eds.), *Form-meaning Connections in Second Language Acquisition* pp: 77-96. London: Lawrence Erlbaum Associates.
[7]   Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition, *IRAL, 41*, 87-106
[8]   Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21,* 181-193.
[9]   Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning; reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second language Instructions,* 258-286.
[10] Hulstijn, J., Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51* (3), 539-558.
[11] Hulstijn, J., Fukkink, R. G. & Simis, A. (2005). Does training in second language word recognition skills affect reading comprehension? An Experimental Study. *The Modern Language Journal, 89*, 54-75.
[12] Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21* (2), 181-193.
[13] Hunt, A & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, Volume 17(1), April, 2005. Retrieved from http://wenku.baidu.com/ view/70c2c70c52ea551810a687da.html.
[14] Joe, A. (1995). Text-based Tasks and incidental vocabulary learning. *Second Language Research, 11*, 149-58.
[15] 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.
[16] Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58* (2), 285-325.
[17] Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *61*(I), 100-140.
[18] Laufer, B. (2001). Reading, Word-Focused Activities and Incidental Vocabulary Acquisition in a Language. *Prospect, 16(3),* 44–54.
[19] Laufer, B. (2004). Focus on form in second language vocabulary learning. The 14th EUROSLA Conference, San Sebastián, 8-11 September 2004.
[20] Morra, S., & Camba, R. (2009). Vocabulary learning in primary school children: Working memory and long-term memory components. *Journal of Experimental Child Psychology*, *104*,156-178.
[21] Nakata, T. (2008). English vocabulary learning with word lists, word cards, and computers: Implications from cognitive psychology research for optimal space learning. *ReCALL, 20(1),* 2-30.
[22] Nation, I. (2001). Learning vocabulary in another language (p.232). Cambridge: Cambridge University Press.
[23] Newton, J 1995. 'Task-based interaction and incidental vocabulary learning: A case study'. *Second Language Research, 11*, 159–77
[24] Paribakht, T.S., & Wesche, M. (1999). Reading and incidental L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in second Language Acquisition, 21*, 195-224.
[25] Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber Walkenfeld & Hernstast (1991), and Knowlton & Squire (1996) and relevance for SLA. *SSLA, 27,* 235-268.
[26] Rosa, E., & Leow, R. P. (2004). Awareness, different language conditions, and second language development. *Applied Psycholinguistics, 25*, 269-292.
[27] Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycle on lexical acquisition and retention. *Language Teaching Research, 6(3)*, 183-222.
[28] San Mateo Valdehita, A. (2003, 2004). Learning from lexicon in Spanish as second language. Unpublished M.Phil, thesis, UNED. June, 2005. Retrieved from http://www.mec.es/redele/biblioteca2005/san_mateo.shtml.

[29] Sanchez, M. J. (2004). Effect of instruction with expert patterns on the lexical Learning of English as a foreign language. *System, 32(1)*, 89-102.

[30] Schmitt, N. (2008). Instructed language vocabulary learning. *Language Teaching Research, 12(3)*, 329-363.

[31] Stoller, F. L. & Grabe, W. (1993). Implications for L2 vocabulary acquisition and instruction from L1 vocabulary research. In T. Huckin, M. Haymes, & J. Coady (Ed.), *Second language reading and vocabulary learning* (pp. 24-45). Norwood, Nj: Ablex.

[32] Stringer, D., & Bardovi-Harling, K. (2010). Variables in second language attrition: Advancing the state of the art. *Studies in Second Language Acquisition, 32 (1)*, 1-45.

[33] Webb, S. (2005). Receptive and productive vocabulary learning. The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 17*, 33-52.

[34] Zimmerman, C. (1997). Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly, 31,*121-140.

Elaheh Touti received her M.A. in TEFL from the faculty of Persian Literature and Foreign Languages, Islamic Azad University, South Tehran Branch, Iran in 2012. She has taught English at different levels to Iranian EFL Learners for more than 10 years in several English institutes and she, currently, is teaching general and academic English courses at the university of Applied Science and Technology in Tehran, Iran. She has also translated two of the Mc Grow Hill books, including 101 American English Idioms and 101 American English Riddles into L1 published by Sokhanvaran Publications, Tehran: Iran, 2016. Her research interests are second language vocabulary instruction and effective methods of lexical enhancement, EFL writing, task-based language instruction and psycholinguistics.

Roohollah Maleki was born in Nahavand, Iran in 1980. He holds MA in TEFL from Islamic Azad University, South Tehran Branch. He teaches in the English Center at Shahid Sattari Aeronautical University of Technology and Science, Tehran, Iran. His research interests are second and third language acquisition, contrastive analysis, discourse analysis, testing and material development. He has published several articles in language learning acquisitions e.g., A Comparative Study of the Attribution for Failure of English Majors Students versus that of their Instructors published by International Journal of Language Learning and Applied Linguistics World in April 2014.