Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi
Mustafa Kemal University Journal of Graduate School of Social Sciences
Yıl/Year: 2016 ♦ Cilt/Volume: 13 ♦ Sayı/Issue: 35, s. 77-92

THE EFFECT OF TASK-INDUCED INVOLVEMENT LOAD ON INCIDENTAL VOCABULARY ACQUISITION

Tuncay KARALIK
Anadolu Üniversitesi, Eğitim Fakültesi, tuncaykaralik@anadolu.edu.tr

Ali MERÇ
Anadolu üniversitesi, Eğitim Fakültesi, amerc@anadolu.edu.tr

Abstract

This study aimed to investigate the effect of task-induced involvement load (Laufer & Hulstijn, 2001) on incidental vocabulary gain and retention. A total of 139 first year ELT students from eight intact classes taking Academic Reading Course in the fall semester of 2015-2016 were assigned to four experimental groups randomly: fill-in with glossary, fill-in by searching, retelling with glossary and retelling by searching. Two different texts were designed for each task. Each text was accompanied by reading comprehension questions and contained ten target words. The groups were to complete the reading comprehension tasks and the vocabulary tasks they were assigned to. The fill-in groups completed the gaps in the text with the target words with or without glossary. Retelling groups were to retell the reading texts incorporating the target words with or without glossary. After the completion of the tasks, an unannounced post-test was administered. Two weeks later, an unannounced delayed post-test was also given.

The comparison of the groups on immediate and delayed post-tests using one-way ANOVAs showed that the tasks with higher level of involvement load yielded higher vocabulary gain and retention. However, the only significant differences were between retelling by searching and fill-in groups on the delayed post-test, which provided partial support for the Task-induced involvement load hypothesis.

Keywords: *Task-induced involvement load hypothesis, Vocabulary gain, Vocabulary retention, Incidental learning*

GÖREV KAYNAKLI KATILIM YÜKÜNÜN RASTLANTISAL KELİME ÖĞRENMEYE ETKİSİ

Özet

Bu çalışma Laufer ve Hulstijn (2001) tarafından önerilen Görev Kaynaklı Katılım Yükünün rastlantısal kelime öğrenimi ve kalıcılık üzerindeki etkisini bulmayı amaçlamıştır. 2015-2016 Güz yarıyılında Akademik Okuma dersine kayıtlı toplam sekiz şubeden 139 İngilizce Öğretmenliği birinci sınıf öğrencisi dört deney grubuna atanmıştır: sözlükçe yardımıyla boşluk doldurma, sözlük kullanarak boşluk doldurma, sözlükçe yardımıyla tekrardan yazma ve sözlük kullanarak tekrardan yazma. İki farklı okuma parçası dört kelime etkinliğine göre hazırlanmıştır. Okuma parçalarından her biri 10 hedef kelime ve okuma anlama bölümleri içermektedir. Katılımcılar okuma anlama bölümlerini ve atandıkları etkinlikleri tamamlamaktan sorumlu tutulmuştur. Boşluk doldurma grupları, okuma parçasında yer alan boşlukları hedef kelimeleri kullanarak sözlükçe yardımıyla veya sözlük

kullanarak doldururken tekrardan yazma grupları hedef kelimeleri kullanarak okuma parçasının kendi ifadeleri ile sözlükçe yardımlı veya sözlük kullanarak yazmıştır. Kelime etkinlikleri ve okuma anlama bölümlerinin tamamlanmasından hemen sonra bir son test uygulanmıştır. Uygulamadan iki hafta sonra ise bir kalıcılık testi verilmiştir.

Kelime etkinliği gruplarının tek yönlü ANOVA yöntemi kullanılarak karşılaştırılmaları sonucunda daha yüksek katılım yüküne sahip etkinlik gruplarının daha fazla kelime öğrendikleri ve daha yüksek kalıcılık sağladıkları görülmüştür. Fakat sadece sözlük kullanarak tekrardan yazma ve boşluk doldurma grupları arasında anlamlı fark bulunmuştur. Bu yönüyle bu çalışma sonucunda görev kaynaklı katılım yükü kuramsal çerçevesine kısmi destek bulunmuştur.

Anahtar kelimeler: *Görev kaynaklı katılım yükü hipotezi, Kelime öğrenimi, Kalıcılık, Rastlantısal öğrenme*

Introduction

The role of vocabulary knowledge in determining success in language learning has led researchers to investigate ways to foster vocabulary expansion. To this end two main types of vocabulary learning have been acknowledged within the literature of vocabulary acquisition in a second/foreign language: intentional learning and incidental learning. Intentional vocabulary learning has been referred to as paying deliberate attention to commit the lexical items to long term memory by applying some retention strategies such as rehearsals and mnemonics for preparing a later recall test (Hulstijn, 2003; Schmidt, 1984). Incidental learning, on the other hand, is defined as learning vocabulary as by-product of focusing on meaning (e.g., reading and completing a task) by Laufer and Hulstijn (2001). Regarding the value of each type of learning, not all of the vocabulary language learners possess can be attributed to intentional attempts; rather most of the vocabulary stock a learner has is acquired incidentally (Nagy & Herman, 1987).

However, as Paribakht and Wesche (1997) suggested, it is not always possible that learners will acquire all the words they encounter incidentally and teachers need to draw their students' attention to the words they want their students to learn. They suggest that teachers can achieve this with the help of instructional interventions (e.g., vocabulary tasks accompanying reading texts). Studies regarding the effect of vocabulary tasks in promoting incidental vocabulary learning suggested that participants acquire more words when they complete vocabulary tasks than when they don't engage in any tasks. Therefore, preparing and selecting effective tasks for incidental vocabulary acquisition are important concerns for instructors.

Following the suggestion Nagy and Herman (1987) and Paribakht and Wesche (1997) made, there have been attempts to determine the factors contributing to incidental acquisition of vocabulary. As a result of these attempts, incidental vocabulary learning through tasks was found to be dependent on the frequency of encounters with a word, dictionary use and processing depth. In order

to conceptualise a framework that accounts for the depth of processing, Laufer and Hulstijn (2001) proposed Task-induced Involvement Load Hypothesis (TILH) consisting of evaluation, need and search components.

Evaluation means the judgement of the learners about whether a given word fits into a specific context or their choice of the correct homonym of a word to match a sentence or a piece of discourse. Laufer and Hulstijn (2001) described the component as 'selective decision based on a criterion of semantic and formal appropriateness (fit) of the word into its context' (p. 15). The component of evaluation can exist in a task at three levels: moderate and strong or absent. Laufer and Hulstijn (2001) demonstrated how the level of evaluation could be adjusted with the following tasks: a fill-in the blanks task requiring the learner to select the appropriate word from a list of words and another task requiring the learner to use a word in meaningful sentences. The first task will pose moderate evaluation as the context where the word will be used is limited to just the list of words and the learner just needs to decide which word will suit best to the gap. However, writing an original sentence using the word will require strong evaluation as the learner needs to make some syntactic and semantic decisions such as what kinds of words can come before and after the target words (TW) and what kind of a sentence the word should be used in.

Search component is related to whether the learners need to find out the meaning of the given words himself by using any source like dictionaries or software. If the meaning of a specific word is looked up in a dictionary or in any other sources, then there exists search. However, when the definitions of the words are provided in a dictionary or the meanings can be inferred from the context, then, the 'search' component is not present.

Need is the motivational component of the construct of involvement load and it is related to why a specific word has to be known. The component is operationalised at three levels: moderate, strong or it can be absent. Whether the drive to use the word is externally or internally imposed determines the degree of need in an activity. If a word is required by an external factor such as the teacher's demand or the requirement of a task, then need will be moderate. On the other hand, if the learner wants to use the word to serve his own purposes such as talking about something or describing something, need will be strong. If the word is not needed in a specific context, then need will be absent.

The TILH by Laufer and Hulstijn (2001) posits that existence or absence of these three components determines the overall Involvement Load Level (ILL) of a given task. The existence of the components is graded through minuses (-) and plusses (+). If a component does not exist in a task, it is given – and counted as 0. However, when a task includes a component at moderate level that component is given + and counted as 1. When a component exists in a task at strong level, it is given ++ and counted as 2. Adjusting the existence level of the components would permit creating different tasks with different levels of involvement load. It is

assumed that the higher the involvement load of a task is, the better the task will contribute to vocabulary gain and retention.

There have been several studies since the proposition of the framework. The first one of these studies was Hulstijn and Laufer (2001) in which the level of evaluation component was differentiated across reading comprehension with marginal glosses, reading comprehension plus gap fill and composition writing incorporating the TWs. The first task induced moderate need as the knowledge of the TWs was obliged by the comprehension questions. The second task induced moderate need and moderate evaluation. The third task was loaded with moderate need and strong evaluation. Therefore, total involvement load indexes of the tasks were 1, 2 and 3 respectively. The study took place in Israel and Netherlands simultaneously. The post-tests required the participants to provide L1 equivalents or English explanations of the TWs. Each correct answer was counted as 1 and the wrong answers were counted as 0. The groups then were compared on their immediate and delayed post-test scores. Data analysis showed that composition writing group significantly outscored the other groups on both post-tests in both countries. As for gap-filling and reading comprehension with glossary, there was a significant difference between these groups on both immediate and delayed post-tests in Israel. However, the groups in Netherlands did not differ from each other on the post-tests. The study provided strong empirical support for the hypothesis that tasks with higher involvement load level (ILL) yield higher incidental vocabulary acquisition.

Tu (2004) prepared different tasks containing varying ILLs and explored the effect of tasks on incidental vocabulary acquisition. The tasks were reading comprehension with marginal glosses (moderate need as the comprehension questions required the participants to refer to the TWs), reading comprehension plus fill-in (inducing moderate need and moderate evaluation) and writing a composition incorporating the TWs (inducing moderate need and strong evaluation) with involvement TILLs of 1, 2 and 3 respectively. The groups were compared on their immediate and delayed post-test mean scores. The results showed that composition writing group significantly outperformed the other groups on both post-tests. Similarly, the second highest loaded task, reading comprehension plus gap-fill, yielded significantly better results than reading comprehension only task. The results of the study strongly supported involvement load hypothesis.

The significant effect of increasing the presence of evaluation was proven in other studies as well. The results suggested that tasks with higher ILLs were found to be better in promoting incidental vocabulary acquisition (Behbahani, Pourdana, Maleki & Javanbakht, 2011; Feng, 2015; Sarbazi, 2014).

In addition to studies testing the adjustment of evaluation component only, there have been studies testing the effect of search and evaluation components as well. Sarani, Negari and Ghaviniat (2013) designed six tasks with

varying involvement loads. Three of the tasks were receptive and the other three were productive. Such a design would help investigating whether different tasks (receptive, productive) with similar involvement load levels would yield similar results in terms of incidental VG and VR. Receptive tasks were true-false (moderate need), matching (moderate need and moderate evaluation) and multiple choice (moderate need, search and moderate evaluation) inducing involvement load indexes of 1, 2, and 3 respectively. The distribution of the components in the receptive set was designed in such a way that the researcher would be able to see if adding a component is really effective (by adding evaluation in the second task and by adding search in the third task). Similarly tasks in the productive set induced similar levels of involvement load and these tasks were short response (moderate need), fill-in (moderate need and moderate evaluation) and sentence writing (moderate need and strong evaluation). First, the receptive and productive sets were analysed separately. The analysis on the receptive tasks showed that multiple choice group with the highest involvement load significantly outperformed the other groups. Similarly there was a significant difference between matching and true false tasks on both immediate and delayed post-tests. This result provided strong support for the assumption that increasing the ILL, whether by adjusting search or evaluation, will increase vocabulary retention as well. Another finding supporting this assumption was that there was no difference between sentence making and multiple choice groups which contained different distributions of evaluation and search. This result indicated that however the components are distributed, as long as the total ILL is equal there will be no differences in vocabulary retention.

Using a similar design to that Sarani et al. (2013) used; Ghabanchi, Davoudi and Eskandari (2012), Hazrat (2015) and Pourakbari and Biria (2015) also found similar results proving the significant effect of ILL on vocabulary acquisition.

On the other hand, there were also contradictory results which suggested that increasing the level of evaluation (Bao, 2015; Beal, 2007; Jahangiri & Abilipour, 2014; Keyvanfar & Badraghi, 2011), and inserting search (Haratmeh, 2012; Jahangard, 2014; Marmol & Sanchez-Lafuente, 2013) didn't always result in increased vocabulary gain. Participant related factors such as attention span, writing skills and dictionary use skills were found to be hindering the effect of ILL. Therefore, this study intended to verify the effect of ILL of a task on vocabulary acquisition using a group of participants with a higher proficiency level to control for learner-related factors. Using the framework of TILH, the present study aimed to test the effect of different tasks with varying and equal ILLs on vocabulary acquisition among English Language Teaching (ELT) students. The questions to be answered at the end of the study were:

1) Do different vocabulary tasks with varying levels of involvement load differ in terms of their contribution to vocabulary gain and retention?

2) Do different vocabulary tasks with similar levels of involvement load operationalised by different components (search and evaluation) lead to similar amounts of vocabulary gain and retention?

Methodology

Participants

The subjects of the present study were first-year students enrolled in the English Language Teaching program at Anadolu University. The study employed two reading texts implemented at different times, and therefore, the number of the participants attending each implementation was different. A total of 131 participants were included in the data analysis of the first implementation while 139 participants were present in the second implementation. The number of the participants assigned to different tasks and across two implementations is presented in Table 1 below.

Table 1: *Number of the Participants across Tasks and Implementations*

	First Implementation	Second Implementation
Fill-in with Glossary	30	33
Fill-in by Searching	39	39
Retelling with Glossary	30	33
Retelling by Searching	32	34
Total	131	139

The age range of the participants was 17-21. The participants were from eight intact classes enrolled in the Academic Reading course.

The proficiency levels of the subjects participating in the study were upper intermediate and above as either they had just passed the preparation school examination of the institution, or they were already exempt from preparatory education or they had completed the preparatory education. When the participants were compared in terms their proficiency scores on the preparation school examination conducted at the beginning of the academic calendar, the classes were found to be similar to each other. Therefore, the comparison allowed the researcher to assume that the groups were homogeneous and proficiency differences would not interfere with the overall results of the study.

Instruments

Several instruments consisting of two reading texts, vocabulary tests (pre-test, immediate post-test, delayed post-test) and vocabulary tasks were used in the study.

Reading texts were taken from an IELTS practice web site. The two reading texts contained 10 TWs (all nouns) each and were followed by reading comprehension sections. The TWs were aggression, equilibrium, indifference, kinship, morality, offspring, propensity, respiration, solidarity and wit for the first text. The TWs in the second text were acumen, apprentice, audit, deficiency, dilemma, distress, entrepreneur, frustration, gratification, and recruitment.

Vocabulary tests required the participants to provide the L1 equivalents or English explanations of the TWs. Each correct answer was scored 1 and incorrect ones were calculated as 0. Partially correct answers were awarded 0.5. Vocabulary pre-test was used to determine whether the participants knew any of the target words. Immediate vocabulary post-test was administered to test initial vocabulary gain. The delayed post-test was conducted to measure vocabulary retention of the participants.

The vocabulary tasks (VT) designed to operationalise varying levels of involvement load were fill-in the blanks by searching (FBS), fill in the blanks with glossary (FWG), retelling by searching (RBS) and retelling with glossary (RWG). The distribution of the components and the overall ILLs of the VTs are shown in Table 2.

Table 2: *Vocabulary Tasks and Involvement Load Levels*

	By searching		With glossary	
	Fill-in	Retelling	Fill-in	Retelling
Evaluation	+	++	+	+ +
Search	+	+	-	-
Need	+	+	+	+
Total TILL	3	4	2	3

The participants assigned to fill in the blanks by searching group were required to complete the blanks of a gapped version of the reading passage by choosing the appropriate words from a word list, including the TWs and an additional word functioning as a distractor. The participants were encouraged to consult a dictionary or their smart phones for looking up the meanings of the words in the word list.

Fill in the blanks with glossary group was also required to complete the reading passage by filling in the gaps by using the TWs. This time, unlike the fill-in by searching group, both Turkish equivalents and English explanations of the TWs were provided by the researcher by means of a glossary.

Retelling by searching group was required to complete the reading comprehension questions and retell the text by incorporating the TWs encountered in the reading text. The TWs were written in bold to draw the participants'

attention to the words and the context they were used in. Additionally, they were allowed to look up the meanings of the TWs in a dictionary or any other source. The instruction for this task informed the students about what information they had to include in their own versions of the texts.

Quite similar to the retelling by searching group, retelling with glossary group was also required to complete the reading comprehension questions and retell the reading text by incorporating the TWs according to the instructions. However, the participants in this group did not have to consult any sources for the meanings of the words as they were provided with marginal glosses of the words.

Procedures

Prior to the actual study, the materials were piloted with a parallel group of participants no included in the actual study. The purpose of the piloting was twofold: to determine whether the instructions were clear and to discover how much time the tasks required.

Before implementing the tasks, the participants' consent was taken. The actual study adopted a between-subjects quasi-experimental design (Creswell, 1994) without a control group: four different tasks with varying involvement load levels were assigned to four groups from eight intact classes. The implementation process lasted for four weeks in total. The first and the second weeks were devoted to the vocabulary pre-tests, reading comprehension, vocabulary tasks and immediate vocabulary post-tests of the two reading texts. Third and fourth week were devoted to delayed post-tests.

Prior to the implementation of the reading texts, the vocabulary pre-test of the related reading text was administered to check whether the participants knew the meanings of the TWs. The participants were required to provide Turkish equivalents or English explanations of the TWs appearing in an alphabetical order.

After the completion of the pre-tests, the participants were handed out different versions of the texts and vocabulary tasks according to their respective groups. As explained above, different groups completed a different vocabulary task in the allotted time. Once the reading comprehension sections and vocabulary tasks were completed, the worksheets were collected and an unannounced immediate post-test was administered to measure the immediate gains.

A two-week period was set for administering the unannounced delayed post-test to measure the effect of ILL on vocabulary retention. The implementation procedure was the same for both of the reading texts.

Data Analysis

The groups processing the target words at different TILLs were compared on their immediate and delayed post-test scores to answer the first research question regarding the effect of different TILLs on incidental VG and VR. For the second research question, on the other hand, the groups completing tasks with

similar TILLs were compared on both post-tests. Since there were four groups, One-way ANOVA was selected as the statistical analysis method.

Results and Discussion

The different tasks with varying ILLs were compared on their immediate and delayed post-tests across two implementations. The vocabulary scores of the participants in the two implementations are shown in Table 3.

Table 3: *Vocabulary Mean Scores*

	Implementation		Implementation 2	
	Immediate Post-test	Delayed Post-test	Immediate Post-test	Delayed Post-test
Fill-in with Glossary	8,4	5,233	8,758	4,742
Fill-in by Searching	8,513	5,385	8,846	5,346
Retelling with Glossary	8,6	5,633	9,076	5,576
Retelling by Searching	9,172	6,406	9,294	6,309

For immediate post-test results of the first implementation, as Table 3 presents, all the groups, regardless of their different ILLs, gained the meanings of nearly all of the ten target words. The tasks actually led to substantial amounts of incidental vocabulary gain. It is also evident that the lowest mean score was obtained in the fill-in with glossary (FWG) group with an ILL of 2 (M=8.4, SD=0.968), followed by fill-in by searching (FBS) inducing an ILL of 3 (M=8.513, SD=1.574) and retelling with glossary (RWG) containing an ILL of 3 (M=8.6, SD=1.493). The highest mean score was that of retelling by searching (RBS) (M=9.172, SD=0.929) and total ILL for this task was 4. It was found that slight differences existed among the groups on immediate recall of the words, with the higher loaded groups obtaining higher scores. One-way ANOVA conducted to see if these differences were significant showed that the groups didn't differ significantly from each other (F(3,127)=2.258, p>.05).

As for the delayed post-tests of the first implementation, it was found that there was a similar pattern in the line-up of the scores with the lowest mean score belonging to FWG (M=5.233, SD=1.628) group. The highest scoring group was RBS (M=6.406, SD=1.34) followed by RWG (M=5.633, SD=2.046) and FBS (M=5.385, SD=1.411) respectively. The results of the one-way ANOVA indicated that there was a significant difference among the groups in terms of their vocabulary retention scores (F(3,127)=3.371, p<.05). In order to detect which groups differed from one another significantly, a Tamhane's T2 post-hoc test was conducted. As the results of the analysis demonstrated, RBS group (M=6.406, SD=1.34) was found to be

significantly outperforming FBS (*M*=5.385, *SD*=1.411) and FWG (*M*=5.233, *SD*=1.628) groups with a moderate task effect size.

The same analyses were conducted on the second implementation. As for the immediate post-test results of the second implementation, the groups were in the order parallel to their respective ILLs and it was found that the highest mean score was achieved by the RBS group (*M*=9.294, *SD*=0.88) completing the highest loaded task followed by RWG (*M*=9.076, *SD*=1.146) and FBS (*M*=8.846, *SD*=1.107) groups respectively. The lowest immediate mean score was obtained by the participants in FWG group (*M*=8.758, *SD*=1.305) completing the least involving task. As in the first implementation however, these differences were not statistically significant ($F(3,135)=1.594$, $p>.05$).

The delayed post-test comparisons yielded that the highest score was obtained by the RBS group (*M*=6.309, *SD*=1.231) which was followed by RWG (*M*=5.576, *SD*=1.193) and FBS (*M*=5.346, *SD*=1.483) groups. The lowest scoring group was FWG (*M*=4.742, *SD*=2.253). The one-way ANOVA demonstrated that The results of one-way ANOVA showed that the participants completing different vocabulary tasks with different TILLs differed from each other significantly ($F(3,135)=5.569$, $p<.05$) in terms of their vocabulary retention scores. In order to detect where the significant difference was, Tamhane's T2 post-hoc test was exploited and the results of the statistical analysis demonstrated that the participants who completed RBS task (*M*=6.309, *SD*=1.231) significantly outperformed their peers in the FBS (*M*=5.346, *SD*=1.483) and FWG groups (*M*=4.742, *SD*=2.253) with a moderate task effect size.

Equally loaded tasks were also compared over two implementations. The comparison of all the groups on the immediate post-test of the first implementation showed that there was not any significant difference between any of the groups ($F(3,127)=2.258$, $p>.05$), which meant that the equally loaded groups didn't differ from each other either with FBS (*M*=8.513, *SD*=1.574) and RWG (*M*=8.6, *SD*=1.493). Similarly, ANOVA results on the delayed post-test of the first implementation showed that the two groups didn't obtain significantly different scores from each other with RWG (*M*=5.633, *SD*=2.046) and FBS (*M*=5.385, *SD*=1.411).

Results of the second implementation as well demonstrated that equally loaded tasks lead to similar amount of word gain and retention. The immediate post-test comparison of the scores showed that none of the four groups differed from the others significantly ($F(3,135)=1.594$, $p>.05$), which meant that the equally loaded tasks as well produced similar results with FBS (*M*=8.513, *SD*=1.574) and RWG (*M*=8.6, *SD*=1.493). Delayed post-test scores of the two groups were also similar with FBS (*M*=5.385, *SD*=1.411) and RWG (*M*=5.633, *SD*=2.046).

The results yielded important findings with regard to the construct of involvement load. Most important of them is that, the results obtained from two different implementations conducted at different times demonstrated that the

effect of ILL was consistent over different texts and different target lexical items. Regarding this finding, it can be argued that the construct of involvement load has a great deal of predictive value when learner-related factors are kept under control. Implementing different texts with different target words on the same participants by sticking to similar procedures enabled the construct to successfully predict the potential benefits of different tasks.

The results regarding the immediate post-test comparison of the groups indicated that although higher loaded tasks brought about slightly higher mean scores, these differences were not significant. This is not a result expected by the TILH which assumes that regardless of the components being increased, tasks with higher ILLs will lead to significant differences. However, all the participants regardless of their groups were already able to retain nearly all the targeted words, which might have prevented the groups from differing significantly from each other. As Kim (2008) suggested, the effects of tasks with different involvement load levels may not necessarily be salient immediately but rather become evident over time. However, it can still be argued that counterproductive effects of tasks with higher level of evaluation present in Bao (2015), Beal (2007), Jahangiri and Abilipour (2014) and Keyvanfar and Badraghi (2011) didn't exist in the present study as learner-related factors were kept under control. The participants had a certain level of proficiency in writing compositions and the ones who didn't write a well-developed composition were excluded from the study and these facts may have allowed the evaluation component to take effect. Similarly, contrary to the findings of Haratmeh (2012), Jahangard (2014), Li (2014), and Marmol and Sanchez-Lafuente (2013) who suggested that tasks requiring dictionary use were counterproductive due to learner-related factors such as lack of dictionary use experience, the results of this study indicated that groups doing the same task with search inserted outperformed their peers who were provided with the glossary of the words.

Compatible with the prior interpretation that the groups may not differ from each other significantly on the short-term, the delayed post-test scores were highly contingent upon the ILLs of different tasks. The significant benefit achieved when the search and strong evaluation existed in a task at the same time showed that search and evaluation components had to exist at the same time in a task to produce significant retention of words. Worded differently, as Tu (2004) also suggested, probably because of the relative difficulty of the target words, the retention of these words required both dictionary look-ups and text creation, which presents the complementary relationship between the search and evaluation components. Based on the findings, it can be suggested that difficult words such as academic words and false cognates can be taught through tasks which induce search and strong evaluation at the same time.

Another important finding, which this study aimed at uncovering was that equally loaded tasks were found to be yielding similar results on both immediate

and delayed post-tests regardless of the distribution of search and evaluation components. The assumption of the TILH is that regardless of the mental efforts or the time different tasks require, it is the total ILL that will determine the overall vocabulary gain and retention over time. This was the case with the gap filling by searching and retelling with glossary tasks in this study. Although the latter required more time and mental effort to complete, the scores of the two groups on both post-tests were similar, which may have some useful implications for classroom practice. Contrary to the expectations of Laufer and Hulstijn (2001) that search may not be as effective as evaluation component, the search component was found to be substantially affecting vocabulary acquisition when accompanied by strong evaluation. Neither search nor strong evaluation was enough on their own to bring about significantly higher retention of the TWs, but when these two components existed in a task together, they acted as complements of each other. Therefore, it might be better to try to find out how evaluation and search components contribute to vocabulary acquisition through different tasks than trying to compare which component is superior to the other.

All in all, the present study provided strong support for the involvement load construct of Laufer and Hulstijn (2001). Increasing the overall ILL was found to be effective in promoting vocabulary acquisition through tasks. Additionally, controlling for learner-related deficiencies such as dictionary use habits, writing skills and attention span which were present in Haratmeh (2012), Van Polen (2014), and Walsh (2009) was found useful for the components of search and evaluation to take effect.

Implications

The study found that even if the learners were unaware of the fact that they were to be tested on the TWs later, they incidentally acquired and retained most of the TWs by completing the tasks they were assigned to. Therefore, involving the learners in tasks where they are to use the words to accomplish some tasks can be a useful tool for drawing their attention to important words thus promoting incidental vocabulary gain and retention.

Additionally, the results showed that regardless of the components adjusted, the higher ILLs were found to be associated with higher vocabulary gains. This suggests that all the components in the construct carry the same weight. However, the interesting finding that only the retelling by searching task, which induced search and strong evaluation at the same time, yielded significantly higher vocabulary retention may indicate a strong interaction between look-ups and creating a text. When either one of these components of search and strong evaluation is absent, significance cannot be achieved. These components, when they were present in a task alone, did not bring about significantly higher vocabulary retention but when they appeared in a task together, they were effective in promoting vocabulary retention. As this finding suggests, learners will benefit from vocabulary tasks with higher ILLs more when compared to those with

lower ILLs. Designing tasks which include higher levels of evaluation and search components will help learners gain and retain more words. When learners are involved in searching the meanings of the TWs themselves and use these words to create a composition, they will be better at retaining these words. However, drawing on the results of the former studies with lower level participants, the relative effectiveness of involvement load will depend on the learner characteristics such as attention span, writing skills, vocabulary knowledge and dictionary use habits. These factors should be taken into consideration before implementing vocabulary tasks.

Another suggestion which can be useful for classroom practices is that two tasks with equal ILLs result in similar vocabulary gain and retention when the requirements of the tasks are fully met. The study found that a gap fill task by using a dictionary which required 40 minutes, and a retelling task with the help of a glossary lasting for 50 minutes respectively, were not significantly different from each other in terms of their contribution to immediate recall of vocabulary and retention of these words over a two-week period. Therefore, teachers can prefer to use the best task based on their students' writing skills. If their students are not good at writing compositions, they can choose to introduce dictionary use (search) in order to compensate for the absence of strong evaluation.

The results are also beneficial for material designers while designing reading activities. The involvement load construct can be taken into consideration while preparing reading texts and accompanying tasks for specific audiences.

Limitations and Suggestions for Future Studies

The most obvious limitation is that only retelling and gap filling tasks were compared in this study. Using the framework, many other vocabulary tasks can be designed and put under investigation. The present results may be generalised for the tasks under investigation, but further studies are needed for further verification of the framework by means of different tasks.

Similarly, focusing on ELT students, who have a certain level of proficiency and capabilities in using dictionaries and writing skills, the results of the current study may not be generalised for lower proficiency learners. In order to investigate the effect of task-induced involvement load on incidental vocabulary acquisition more thoroughly, further studies can focus on different participants with different proficiency levels.

Moreover, for eliminating the effect of word category on acquisition of the target words, only nouns were put under investigation in the current study. This creates the need for further studies investigating whether the involvement load level produces similar findings regarding different word categories.

References

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

Beal, V. (2007). *The weight of involvement load in college level reading and vocabulary tasks* (Doctoral dissertation). Retrieved from Concordia University CLUES Library Catalogue (LE 3 C66E38M 2007 B43).

Behbahani, S. M. K., Pourdana, N., Maleki, M., & Javanbakht, Z. (2011). EFL task-induced involvement and incidental vocabulary learning: succeeded or surrounded. In *International Conference on Languages, Literature and Linguistics. IPEDR Proceedings, 26*, 323-325.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches.* Sage Publications: Thousand Oaks.

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

Haratmeh, M. S. (2012). Involvement load and task type in task effectiveness: Two aspects of vocabulary knowledge. *International Journal of Academic Research, 4*(4), 86-95. http://dx.doi.org/10.7813/2075-4124.2012/4-4/B.13

Hazrat, M. (2015). The effects of task type and task involvement load on vocabulary learning. *Waikato Journal of Education*, *20*(2), 79-92. http://dx.doi.org/10.15663/wje.v20i2.189

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (Eds.), *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. http://dx.doi.org/10.1111/0023-8333.00164

Jahangard, A. (2014). Task-induced involvement in L2 vocabulary learning: A case for listening comprehension. *Journal of English Language Teaching and Learning*, *5*(12), 43-62.

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

Keyvanfar, A., & Badraghi, A. H. (2011). Revisiting task-induced involvement load and vocabulary enhancement: Insights from the EFL setting of Iran. *Žmogus ir Žodis*, *13*(3), 56-66.

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language learning, 58*(2), 285-325. http://dx.doi.org/10.1111/j.1467-9922.2008.00442.x

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

Li, J. (2014). Effect of task-induced online learning behavior on incidental vocabulary acquisition by Chinese learners: Revisiting involvement load hypothesis. *Theory and Practice in Language Studies*, *4*(7), 1385-1394. http://dx.doi.org/10.4304/tpls.4.7.1385-1394

Mármol, G. A., & Sánchez-Lafuente, Á. A. (2013). The involvement load hypothesis: the effect on vocabulary learning in primary education. *Revista Española de Lingüística Aplicada*, (26), 11-24.

Nagy, W.E., & Herman, P.A. (1987). Depth and breadth of vocabulary knowledge: Implications for acquisition and instruction. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–35). Hillsdale, NJ: Erlbaum.

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

Pourakbari, A. A., & Biria, R. (2015). Efficacy of task-induced involvement in incidental lexical development of Iranian senior EFL students. *English Language Teaching*, *8*(5), 122-131. http://dx.doi.org/10.5539/elt.v8n5p122

Sarani, A., Negari, G. M., & Ghaviniat, M. (2013). The role of task type in L2 vocabulary acquisition: a case of involvement load hypothesis. *Acta Scientiarum Language and Culture*, *35*(4), 377-386. http://dx.doi.org/10.4025/actascilangcult.v35i4.21135

Sarbazi, M. R. (2014). Involvement load hypothesis: Recalling unfamiliar words meaning by adults across genders. *Procedia-Social and Behavioral Sciences*, *98*, 1686-1692. http://dx.doi.org/10.1016/j.sbspro.2014.03.594

Schmidt, R. W. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11–26.

Tu, H. F., & Su, I. R. (2004). *Effects of task-induced involvement on incidental vocabulary learning in a second language*. (Unpublished master's thesis). National Tsing Hua University, Taiwan.

Van Polen, Z. (2014). *The effects of task-induced involvement load and word exposure frequency on L2 incidental vocabulary learning through reading* (Doctoral dissertation). Universiteit Van Amsterdam, Netherlands.

Walsh, M. I. (2009). *The involvement load hypothesis applied to high school learners in Japan: Measuring the effects of evaluation* (Unpublished master's thesis). Birmingham University, Edgbaston, United Kingdom.