

<http://dx.doi.org/10.18649/jkees.2016.15.1.1>

Investigating the Task Involvement Load Hypothesis in EFL Learners' Vocabulary Learning

Jooeun Hyun^{*}

Chung-Ang University

Jieyoung Kim^{**}

Chung-Ang University

Hyun, Joeeun, & Kim, Jieyoung (2016). Investigating the task involvement load hypothesis in EFL learners' vocabulary learning. *Journal of the Korea English Education Society*, 15(1), 1-22.

This study investigated the effect of different types of tasks on vocabulary learning based on the task involvement load hypothesis. A total of 70 university students participated in the research. Four tasks were used in this study: Task 1 involved reading comprehension, Task 2 gap-filling, Task 3 sentence writing, and Task 4 gap-filling with the aid of a dictionary. Immediate and delayed posttests were administered to measure the participants' short-term learning and long-term retention of the target vocabulary in both reception and production. For the first analysis, an ANOVA was performed to compare performance on Tasks 1 to 3. The results revealed a significant difference among the three tasks in the immediate posttest, and the writing task yielded the highest scores. Thus, the involvement load hypothesis was partially supported. The second analysis used a t-test to compare performance in Tasks 3 and 4. The results showed a significant difference only in the delayed receptive posttest. These two tasks had the same effect on initial learning, but different effects on the retention of vocabulary learning.

[vocabulary learning/task involvement load hypothesis/depth of processing/
cognitive factor/어휘 학습/과제관여도/인지처리 깊이/인지적요인]

* The first author

** The corresponding author

I. INTRODUCTION

Learning vocabulary plays an important role in learning a second or additional language, as knowledge of vocabulary provides a basic foundation for communication. The amount of vocabulary that needs to be learnt depends on the goal of language use. Nation (2011) stated that a person needs 120 words and phrases as survival vocabulary for foreign travel but 6,000 to 9,000 words to read unsimplified text without the help of a dictionary. Dongkwang Shin and Yuh Vicky Chon's (2011) study revealed that the English textbooks used in Korean secondary schools presented 7,430 words, despite the Korean National Curriculum permitting around 3,000 words at the high school level, as it takes a lot of time and effort for learners to learn the words they need.

In this context, researchers and teachers have tried to find effective ways to teach and learn L2 vocabulary. Particularly, the learning of vocabulary through a range of tasks has been investigated by a number of researchers (Folse, 2006; Hulstijn, 2001; Keating, 2008). Such research has focused on the effects of task-induced vocabulary learning, the factors that should be involved in the tasks to make students engage in learning, and how different factors affect vocabulary learning. The task-induced involvement load hypothesis was proposed by Laufer and Hulstijn (2001) in an attempt to investigate such issues. The hypothesis is based on the concept of levels of processing proposed by Craik and Lockhart (1972), according to which there is a hierarchy of processing ranging from shallow to deep levels. Greater depth of processing is associated with higher levels of retention and long-term memory. The involvement load hypothesis proposes that the level of processing is affected by three components: need, search, and evaluation. Each component is indexed with a number 0, 1, or 2, depending on the level of processing required, with the sum of these numbers indicating the involvement load of a task. According to this hypothesis, a task with a higher involvement load index will allow language learners to learn and remember words more effectively.

A number of studies have examined the task involvement load hypothesis (Hulstijn & Laufer, 2001; Joe, 1995, 1998; Keating, 2008; ; Jieyoung Kim, 2015; Youjin Kim, 2008). It remains, however, inconclusive whether task involvement load has an effect on short-term vocabulary learning and long-term retention. If a higher involvement load index leads to a more effective task, then tasks with the highest involvement index of 3 may be the most effective for vocabulary learning. Moreover, the impact of each component on task involvement remains ambiguous. In other words, it is not yet clear whether different types of tasks with the same involvement load (e.g., index 3) would have the same effects on vocabulary

learning. Laufer and Hulstijn (2001) suggested that further study is required to reveal the different weights of components in involvement load in vocabulary learning.

The purpose of this study was to investigate the effects of task-induced involvement on English vocabulary learning and retention. This study also examined the relative contribution of two components, search and evaluation within the levels of processing concept. Although the level of involvement load is affected by three components: need, search, and evaluation, this study would focus on the search and evaluation that were cognitive constructs. The need factor is a motivational one, which would be separated from other two factors in different dimensions.

In this context, the research questions for the present study were the following:

1. Do different indices of task involvement load affect initial vocabulary learning and/or retention?
2. Do the search and evaluation components of task involvement load have the same degree of impact on vocabulary learning in both initial learning and retention tests?

II. LITERATURE REVIEW

1. Task-induced Vocabulary Learning

It has been asserted that extensive reading is an effective approach in vocabulary development. Under the extensive reading approach, students learn vocabulary from the context without any awareness. It is the basic concept of incidental learning (Nation, 2011). Many researchers, however, have argued that learning vocabulary should not rely solely on incidental acquisition, but there should also be a deliberate design to focus on certain forms or skills needed for learning (Cairns, Cowart, & Jablon, 1981; Fraser, 1999; Hulstijn, 2003). In line with this argument, task-induced vocabulary learning is presented.

Task-induced vocabulary learning can be explained under form-focused instruction similar to grammar learning (Laufer & Girsai, 2008). Under the approach of form-focused instruction, it is important that learners' attention should be drawn to lexical items with meaning-based comprehension such as real-world tasks. Although there are fewer form-focused vocabulary learning studies compared to grammar studies, diverse instructional techniques on vocabulary learning such as dictionary use, vocabulary exercises involving comprehension, and writing sentences have been

investigated (Ellis, Tanaka, & Yamazaki, 1994; Knight, 1994; Laufer & Hulstijn, 2001; Luppescu & Day, 1993).

Cairns et al. (1981) suggests that intended lexical items should appear in a context for learners to infer and learn. They argue that learners remember words better if some difficulty in interpretation is given. They assert that decision-making difficulties induce better memory of words. Their study indicated that easy processing of words in highly predictable contexts results in poor memorization of the lexical items. Their study also revealed that when learners take advantage of a linguistic bottom-up interpretation of a context or when they encounter difficulties in an interpretation that requires more attention, they might learn the vocabulary better. It is suggested that the context of these deliberate items needs to be included in a reading activity regarded as a task.

Fraser (1999) also espouses for additional activities to be included with reading for vocabulary learning. He claims that lexical processing strategies while reading induce better retention of words through inferring from context. His study emphasized the importance of consulting as an explicit vocabulary-learning strategy such as using dictionary. Furthermore, he also suggests that word identification based on L1 is an effective way of remembering. It means that when learners retrieve an L1 meaning for the unknown L2 word, they can remember the word better. In addition, a cumulative process with subsequent meetings is suggested for vocabulary learning design. In conclusion, Fraser's study implies that instruction which focuses on developing lexical processing strategies can benefit vocabulary learning.

Laufer and Girsai (2008) investigated the effects of explicit contrastive analysis and translation activities on incidental vocabulary learning. In the contrastive analysis and translation group, learners were given a true-or-false task while reading a text without dictionary. However, whenever they encounter any unknown words, they were required to ask the meaning from the teacher. They found that learners in the contrastive analysis and translation group had a higher level of word retention in comparison to the other groups in the study. This is explained through the concept of "noticing." While completing tasks, learners need to ask for meanings, and to search and evaluate meanings to translate text into L1 so that they can focus on the target words. Based on the results, Laufer and Girsai (2008) suggest that a form-focused activity within communicative tasks is effective in vocabulary learning.

Nevertheless, task-induced vocabulary learning needs to be investigated further in order for incidental vocabulary learning to be conducted more effectively. In addition, it is relevant to the active cognitive processes during learning such as the "noticing" hypothesis. It means that the dimension of mental processing might be a

factor to explain learning vocabulary (Hulstijn, 2003).

2. Depth of Processing

Morris, Bransford, and Franks (1977) proposed the levels of processing hypothesis, which entails that deeper processing encourages deeper memory among language learners. According to Morris et al. (1977), the depth of learners' word processing affects their retrieval ability. Craik and Tulving (1975) explained that, once a word is recognized, it may proceed to further processing, such as elaboration and association. In terms of such activity, they indicated that "depth of processing" implies a greater degree of semantic or cognitive analysis. For example, connecting new information to pre-existing knowledge leads to longer retention of information. Through such processing, learners may take better advantage of word learning opportunities.

Craik and Lockhart (1972) also claimed that information gained through deeper levels of processing is retained for longer periods of time. In other words, depth of processing takes on a significant role in long-term memory retention. This notion is in line with Schmitt (2004), who stated that, when students are involved in deeper processing of a word, they remember it for a longer period of time. In particular, Craik and Lockhart (1972) indicated that the processing required for an orienting task is one of the important features of incidental learning. Providing an appropriate orienting task leads to better performance under incidental conditions. This suggests that the task is key in controlling processing for incidental learning.

On the other hand, Barcroft (2002) argued that human processing capacity is limited, and that learners can focus on only one aspect of vocabulary knowledge. In other words, increased semantic processing can reduce attention to form focus processing. For example, when learners pay attention to word meaning, word form may not be learned as effectively. For this reason, Barcroft (2002) suggested that deeper processing does not lead to better learning unless focus of attention is appropriate.

The literature reveals further issues regarding the levels of processing hypothesis. As stated earlier, Craik and Tulving's (1975) definition of depth of processing was criticized for being unclear. In this regard, Laufer and Hulstijn (2001) also indicated that the constituents of levels of processing were ambiguous. Moreover, they raised the question of how levels of processing may be measured. It has also been suggested that further research is needed to investigate the constituents of the levels of processing, as well as how these may be operationalized for learning.

3. The Involvement Load Hypothesis

Laufer and Hulstijn (2001) proposed the involvement load hypothesis, investigating the effect of tasks on vocabulary learning. Involvement load was explained with respect to the motivational-cognitive construct, specifically information processing, which is crucial for learning (Laufer & Hulstijn, 2001). The hypothesis is based on the assumption that word retention in incidental processing is conditional upon three factors: need, search, and evaluation. These factors constitute the dimensions of task involvement, and the degree of involvement load determines vocabulary learning. Involvement occurs through the operation of a combination of these components, so that the way in which and degree to which they are combined affects the involvement load. In conclusion, Laufer and Hulstijn (2001) stated that the greater the task involvement load, the better the effect on vocabulary learning. For a better understanding of the three factors of task involvement load, see Table 1.

Table 1
Components of Task Involvement Load (from Youjin Kim, 2008)

Component	Need	Search	Evaluation
Feature	motivational, non-cognitive dimension of involvement	cognitive dimension of involvement	cognitive dimension of involvement
Operationalization	need to know words for the task	attempt to find the meaning of unknown words attempt to identify the appropriate L2 form for a particular concept	comparison of a given word with other words comparison of a specific meaning of a word with its other meanings
Levels	absent (-) moderate (+) strong (++)	absent (-) present(+)	absent (-) moderate (+) strong (++)

The first component is *need*, a non-cognitive motivational dimension of involvement concerned with students' perceived need to know the vocabulary in order to complete the task. The level of involvement load for need may be one of three degrees, namely absent (indexed as (-) or 0), moderate (indexed as (+) or 1), or strong (indexed as (++) or 2). For example, when tasks requires the target words, the degree of need might be (+) or the number 1. And when learners feel the need to know words, need is strong and is indexed as (++) or the number 2.

The second component, *search*, is a cognitive dimension of involvement, entailing the attempt to find the meanings or forms of unknown words. In terms of degree,

search may be either present (indexed as (+) or 1) or absent (indexed as (-) or 0). Thus, no word forms or meanings are provided in the task of search (+) so that students are required to ask to someone or find a dictionary to search the unknown words.

The last component is *evaluation*, a cognitive dimension inducing information processing. It is concerned with comparing given words or identifying a word and determining whether or not it fits the context. The level of involvement load of evaluation may be absent (indexed as (-) or 0), moderate (indexed as (+) or 1), or strong (indexed as (++) or 2). In moderate evaluation, learners are required to recognize differences between words, whereas strong evaluation requires them to become involved in decision-making, such as how to combine words or what additional words are required.

With respect to the impact of these components, each is allocated the same impact. However, the different degrees of involvement load of different components or combinations have not yet been researched. Laufer and Hulstijn (2001) also suggested that the relative weight of each component should be established by further research, as the three components may not have equal influence on vocabulary learning.

4. Empirical Research on the Task Involvement Load Hypothesis

Much research has offered support for the task involvement load hypothesis. Hulstijn and Laufer (2001) offered empirical evidence for it by investigating three tasks, namely (1) reading comprehension with glosses, (2) gap-filling reading, and (3) composition. The varying involvement loads of these tasks revealed significantly different effects on vocabulary learning. In immediate tests, as the involvement load became higher, the learners' performance appeared better in terms of their scores. In terms of retention tests, the composition task yielded better outcomes than reading comprehension and gap-filling. However, there was no significant difference between the latter two tasks (Hulstijn & Laufer, 2001).

Keating (2008) also examined task effectiveness for vocabulary learning among low-proficiency learners of Spanish, using the same types of tasks as Hulstijn and Laufer (2001). According to Keating (2008), low-proficiency learners performed remarkably better following the task with a higher involvement load in both receptive and productive tests. In Hulstijn and Laufer's (2001) study, time-on-task was not considered, and this was regarded as a limitation. In contrast, Keating (2008) controlled time-on-task to prevent the effect of different amounts of exposure to the activity.

Youjin Kim (2008) conducted two experiments to investigate the task involvement load hypothesis. The first experiment revealed a significantly different effect of task involvement load in both immediate and delayed posttests. According to Youjin Kim (2008), participants scored significantly higher on a composition task than on a gap-filling task and a reading comprehension task. As there appeared to be no significant difference between the latter two tasks with the same involvement load index, the effect of two tasks with the same involvement load on vocabulary learning was investigated in a second experiment, using a composition task and a sentence writing task. In the composition task, participants wrote an essay of one to three paragraphs about a given topic, and were required to include 10 target words that were listed with their English explanations. The participants in the sentence writing group were required to write an original sentence for each of the same 10 target words, and were provided with the glosses. Both tasks had an involvement load index of 3, with the same task components: need (1) and evaluation (2). The study revealed that the two tasks with identical involvement load had equal effects on both initial learning and vocabulary retention. Therefore, Youjin Kim (2008) suggested that tasks with higher involvement may be more effective for vocabulary learning, and that teachers should use diverse types of tasks with high involvement loads. Although it verified a similar effect of tasks with the same task involvement load, a further issue was also raised, namely whether there might be different degrees of involvement load for each component (need, search, and evaluation).

In the Korean EFL context, there have also been many studies related to the task involvement hypothesis (Jinhee Huh & Byungbin Im, 2008; Jieyoung Kim, 2015; Yeonkyoung Lee & Jieyoung Kim, 2014; Hyunmi Sung, 2013). Jinhee Huh and Byungbin Im (2008) administered four types of tasks, namely reading-only, rote-writing, sentence-context, and sentence writing. Each task had a different level of involvement load, namely 0, 1, 2, and 3, respectively. As predicted, the reading-only group performed significantly worse than any of the others, and the sentence writing task resulted in the highest scores in both immediate and delayed posttests, thus supporting the findings of Hulstijn and Laufer (2001). Based on these results, Jinhee Huh and Byungbin Im (2008) suggested that language teachers and curriculum planners should consider deeper processing tasks with higher involvement load, which may lead to higher gains in vocabulary learning.

Hyunmi Sung (2013) also examined the effect of involvement load on the vocabulary learning of Korean high school learners who varied in their English proficiency levels. The three tasks used in this study were reading comprehension (load index 1), reading and gap-filling (load index 2), and unscrambling sentences (load index 3). The unscrambling of sentences, with an involvement load of 3,

required participants to complete sentences based on a Korean translation by unscrambling the given words. The results showed that the different task involvement loads had an effect only for high-proficiency learners. In this high-proficiency group, mean scores for the reading comprehension and the reading and gap-filling tasks differed significantly in initial learning and long-term retention, but the difference between the reading and gap-filling task and the unscrambling sentences task was not significant. Hyunmi Sung (2013) suggested that Korean high school classrooms should adapt tasks with higher involvement loads to save time and energy for task development.

Yeonkyung Lee and Jieyoung Kim (2014) conducted similar experiment using the materials, reading and comprehension task (load index 1), reading comprehension and gap-filling task (load index 2), and writing sentences (load index 3) as Hulstijn and Laufer (2001). However, they used nonsense words as target words to avoid students' previous knowledge of the target words. They also found that students who conducted writing sentences (need=1, evaluation=2) outperformed on post-test and delayed post-tests. It indicates that the task with the highest involvement load has greater advantage on learning new words as well as retention of the words. Thus, their study validate the task involvement load hypothesis with the use of nonsense words.

Ever since the involvement load hypothesis was proposed by Laufer and Hulstijn (2001), many studies have been conducted to find supporting evidence. Despite this, the relative weight of the different components and the index of involvement load have not yet been discussed. The few studies that have examined the issue remain inconclusive. This study aimed to investigate the relative weight of each component, as well as to provide supportive evidence for the task involvement load hypothesis.

III. METHOD

1. Participants

A total of seventy university students who were sophomores and juniors majoring in English education in Seoul participated in this study. In the initial test, there were 74 participants, but three of these missed the delayed posttest and one student said he knew a target word. Thus, the data of 70 participants were considered in the study. Thirty-nine students out of 70 reported their TOEIC score and their average TOEIC score was 858.

Table 2
The Participant Groups

	Group 1	Group 2	Group 3	Group 4
Sample size	16	17	19	18
Treatment	Task 1	Task 2	Task 3	Task 4

The participants were randomly divided into four groups, and each group completed one task, named Task 1 to Task 4 (see Table 2). They were in the two classroom and distributed each task materials according to the groups. Sixteen participants comprised the Task 1 group, 17 the Task 2 group, 19 the Task 3 group, and 18 the Task 4 group.

2. Procedure

The four groups completed their respective tasks at the same time for twenty minutes. The time-on-task was equalized across the four tasks; only the involvement load varied. Participants were given brief explanations about the tasks prior to beginning the tasks. The paper-based task materials were provided to participants, and they were requested to follow the directions on the paper.

According to Keating (2008), who conducted the similar experiment, all the groups were given 20 minutes to complete the task. After twenty minutes, all groups were asked to submit their task papers. An immediate posttest was conducted immediately after task completion. The posttest consisted of a receptive and a productive part, and the receptive test was distributed first. After the receptive answer sheet was collected, the productive test was provided to the participants. Five minutes were allowed for each test. The time allocation is sufficient to complete the task in the pilot test to decide the target words. Upon completion of the tasks and tests, participants were asked to check if they knew any of the target words. One student checked a word that he knew and his data was excluded from the analysis.

Two weeks later, a delayed posttest was conducted in the same manner as the immediate posttest described above. The format of the test was also identical to the immediate posttest, but the order of the test items was modified.

3. Materials

1) Tasks

The three tasks used by Hulstijn and Laufer (2001), each with a different task

involvement load, were used in the present study, as well as an additional task (see Table 3). A text adapted from the textbook “Reading to Writing,” was used in all tasks. And the activity of each task was modified based on Hulstijn and Laufer (2001). As mentioned above, each task was completed by a different group of participants.

Table 3
Tasks with Varying Involvement Load

	Activity	Involvement Component			Involvement Load index
		Need	Search	Evaluation	
Task 1	Reading Comprehension	1	0	0	1
Task 2	Gap-filling	1	0	1	2
Task 3	Sentence Writing	1	0	2	3
Task 4	Gap-filling with dictionary	1	1	1	3

Task 1 was a reading activity with comprehension questions and ten target words with L1 glosses; its task involvement load index was 1 (need=1). The comprehension questions consisted of five true vs. false and four multiple choice questions. To control time-on-task, a graphic organizing activity and L1 writing summary activity were included.

Task 2 was a gap-filling task. It presented a reading text identical to that in Task 1, with four multiple choice comprehension questions. It also included a gap-filling activity with ten target words deleted, and a list of the words with L1 translations and L2 explanations. Participants read the text and filled in the blanks with appropriate words from the list. The task involvement load index was 2 (need=1, evaluation=1).

Task 3 was a sentence writing task with a task involvement load index of 3 (need=1, evaluation=2). Participants were required to write an original sentence for each of the ten target words. The L1 glosses were given, and each sentence had to consist of more than seven words, including the target word.

Task 4 had the same involvement load as Task 3, but different components. The task was identical to that of the gap-filling in Task 2, but with the target word list provided instead of the L1 glosses. Participants were required to search for the meanings of the target words in a dictionary to determine the appropriate word for each blank. Thus, this task involved the components search (1) and evaluation (1), rather than evaluation (2), as in Task 3. The task involvement load of Task 4 was 3 (need=1, search=1, and evaluation=1), as was the load of Task 3.

2) Target words

Based on previous studies such as Hulstijn and Laufer (2001) and Youjin Kim (2008), a total of ten target words were used in this study: *scour*, *pecuniary*, *travail*, *meritorious*, *predispose*, *demarcate*, *captivate*, *metamorphosis*, *spawn*, and *stampede*. The words were low-frequency synonyms taken from a dictionary to replace original words in the text. To verify the relevance of the target words, whether these target words were unknown vocabulary was tested to learners in another class with the same English proficiency level as the participants prior to conducting the experiment. Thus, it is assumed that all these target words were unknown words for the target participants.

4. Instruments

To measure the participants' vocabulary learning and retention, immediate and delayed posttests were conducted. Both receptive and productive vocabulary knowledge were measured by a word recognition test and a recall test, respectively. According to Nation (2011), receptive vocabulary knowledge involves perceiving the form of a word and productive knowledge involves expressing a meaning through retrieving and producing the form.

The recognition test was a multiple choice test that required participants to choose the correct L1 meaning for each target word from among four words, three of which were distracters. The recall test was a writing test, requiring participants to write down the word equivalent to a given L1 meaning.

5. Analysis

The data collected from the immediate and the delayed posttests comprised the total scores for each test, receptive and productive. One point was assigned per correct item, for a total out of ten points.

SPSS 18 was used to analyze the data. To address the first research question, a one-way ANOVA was used to compare the mean scores for Tasks 1, 2, and 3. Levene's test was conducted to determine homogeneity of variance across the three groups in advance. Scheffe's post hoc comparison was conducted to examine the significant differences across the three tasks. In addition to these analyses, to address the second research question, the scores for Tasks 3 and 4 were analyzed by independent-samples *t*-tests.

IV. RESULTS

1. The Effect of Task Involvement Load

To investigate the effect of task involvement load, the mean scores for Tasks 1, 2, and 3 were analyzed in terms of descriptive statistics and a one-way ANOVA. Table 4 shows the descriptive statistics for the immediate posttest.

Table 4
Descriptive Statistics for the Immediate Posttest

	N	Receptive Test		Productive Test	
		Mean	SD	Mean	SD
Task 1 (1)	16	8.06	2.081	4.06	2.351
Task 2 (2)	17	9.41	.939	5.29	2.544
Task 3 (3)	19	9.53	1.020	7.37	2.650
Total	70	9.06	1.413	5.67	2.775

Note: The involvement load index is indicated in parentheses in column 1.

According to the immediate posttest descriptive statistics, Task 1, of which the involvement load index was 1, yielded the lowest mean scores in both the receptive and productive posttests. The group doing Task 2 achieved the second lowest mean score in both tests, falling between the scores of the Task 1 and Task 3 groups. Moreover, Task 3, which had an involvement load index of 3, resulted in the highest mean scores on both tests. In general, all participants performed better on the receptive than on the productive test.

Prior to comparing the means of the three groups, the homogeneity of variances was examined by Levene's test. According to this test, the significance values were higher than 0.05, so the three groups had equal variances (Sig=.067, Sig=.855), meaning that they may be considered homogeneous.

As indicated in Table 5, the results of the ANOVA revealed significant differences across the three tasks for both the receptive and productive tests ($p=.007$, $p=.001$) in the immediate posttest. Table 6 shows the results of a Scheffe's post hoc comparison for the immediate posttest data.

Table 5
ANOVA for Immediate Posttest

	Receptive Test					Productive Test				
	η^2	df	MS	F	p	η^2	df	MS	F	p
Groups	22.131	2	11.066	5.545	.007	98.554	2	49.277	7.717	.001

Table 6
Scheffe's Post Hoc Comparison for the Immediate Posttest

	Receptive Test			Productive Test		
	MD	SE	Sig.	MD	SE	Sig.
Task 1 (1) vs. Task 2 (2)	-1.349	.462	.030*	-1.232	.880	.383
Task 1 (1) vs. Task 3 (3)	-1.464	.479	.014*	-3.306	.857	.002*
Task 2 (2) vs. Task 3 (3)	-.115	.472	.973	-2.074	.844	.058

* $p < .05$

Significant differences were found between the mean scores of Task 1 and 2 groups, as well as between Task 1 and 3 groups, for the immediate receptive posttest ($p=.030, .014$). For the immediate productive posttest, Tasks 1 and 3 also differed significantly ($p=.002$). In other words, the Task 1 group, with an involvement load index of 1, recalled the target words less well than did the other groups, whose tasks had higher involvement load indices. Furthermore, the Task 1 group produced fewer target words than did the Task 3 group on the immediate posttest. For both receptive and productive immediate posttests, Task 3 resulted in the highest mean scores.

Turning to the delayed posttest, the descriptive statistics are presented in Table 7. These results show that Task 1 yielded the lowest mean score and Task 3 the highest for the receptive test. The Task 2 group scored between the Task 1 and Task 3 groups on the receptive test. On the other hand, for the productive test, Task 2 resulted in the lowest mean score, with Task 1 yielding the second lowest mean score. As for the receptive test, Task 3 yielded the highest mean score.

Table 7
Descriptive Statistics for the Delayed Posttest

	N	Receptive Test		Productive Test	
		Mean	SD	Mean	SD
Task 1 (1)	16	7.13	2.419	4.81	2.738
Task 2 (2)	17	7.71	2.544	4.65	2.621
Task 3 (3)	19	8.79	1.316	5.63	2.948
Total	70	7.80	2.171	5.09	2.775

The mean scores for the delayed posttest were also analyzed by means of ANOVA, following Levene's test to examine the equality of variances among the three groups. The results of the Levene's test revealed a value greater than .05; thus, the three groups may be assumed to be homogeneous. The ANOVA showed no significant differences across the three tasks for either the receptive or productive delayed posttests (see Table 8).

Table 8
ANOVA for the Delayed Posttest

	Receptive Test					Productive Test				
	η^2	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2	<i>df</i>	MS	<i>F</i>	<i>p</i>
Groups	25.255	2	12.628	2.782	.072	10.086	2	5.043	.652	.525

2. The Different Impact of Task Involvement Components

To address the second research question, *t*-tests were conducted to compare mean scores following Tasks 3 and 4. As mentioned above, Task 3 consisted of need (1) and evaluation (2), and Task 4 of need (1), search (1), and evaluation (1). Both tasks had the same involvement load index of 3, but with a different distribution of components. Table 9 presents the descriptive statistics and *t*-test results.

The descriptive statistics indicate that Task 3 yielded higher mean scores than did Task 4 for both the receptive and productive immediate and delayed posttests. According to the *t*-tests, there were no significant differences in the immediate posttest, implying that Tasks 3 and 4 had similar effects on initial vocabulary learning in terms of both receptive and productive knowledge. Thus, despite different task involvement components, these two tasks with same task involvement load, namely a sentence writing task and a gap-filling with dictionary task, were revealed to be equally effective in initial vocabulary learning.

Table 9
The Descriptive Statistics and *t*-test for the Delayed Posttest

		Mean		SD		<i>t</i>	<i>p</i>
		Task 3	Task 4	Task 3	Task 4		
Immediate Posttest	Receptive	9.53	9.11	1.020	1.023	1.236	.225
	Productive	7.37	5.67	2.650	2.657	1.950	.059
Delayed Posttest	Receptive	8.79	7.44	1.316	2.094	2.354	.024*
	Productive	5.63	5.17	2.948	2.895	.484	.632

**p*<.05

On the other hand, there was a significant difference in scores for the receptive delayed posttest (*p*=.024). The mean score of the Task 3 group was 8.79, which was higher than the mean of 7.44 in the Task 4 group, indicating that the sentence writing task was more effective than the gap-filing with dictionary task for receptive vocabulary retention. Thus, despite of same involvement load index of 3, these two tasks affected the retention of new vocabulary differently.

V. Discussion

1. The Effect of Task Involvement Load

To answer the first research question, concerning the effects of task involvement on vocabulary learning and retention, the mean scores for Tasks 1, 2, and 3 were compared. An effect of involvement load was found for initial vocabulary learning but not for vocabulary retention. This finding partially supports the task involvement load hypothesis that learners will learn vocabulary better through tasks with a higher involvement load.

The three tasks differed significantly in terms of mean scores for both the receptive and productive immediate posttests. In particular, the participants performed better in initial vocabulary learning following sentence writing (Task 3), which had the strongest involvement load, than following reading comprehension (Task 1), with the lowest involvement load. This result could provide supportive evidence for Hulstijn and Laufer (2001), who argued that the composition task with the highest involvement load would produce the best outcomes. Deeper processing is invoked when learners engage in tasks with a higher involvement load, allowing better vocabulary learning.

There were, however, no significant differences between the reading comprehension and gap-filling groups for the productive immediate posttest or between the gap-filling and sentence writing groups on either the receptive or productive immediate posttests. The studies of Hulstijn and Laufer (2001) and Youjin Kim (2008) showed similar results.

There were no significant differences in mean scores across groups for the delayed posttest. Thus, no effect of involvement load was revealed for the retention of vocabulary learning. This finding is somewhat contradictory to those of Hulstijn and Laufer (2001), who reported higher retention following Task 3 than following other tasks. Hyungsun Kim and Yoonhee Na's (2010) study, in which they investigated three tasks with different involvement load indices, also found no effect on Korean university students' vocabulary retention. They found that participants performed better following a task with a higher involvement load in the immediate posttest, but no significant differences between tasks in the delayed posttest. Hyunmi Sung (2013) found that high-proficiency high school students performed better with a higher involvement load task. However, unlike in the current study, there were significant differences on both the initial and retention tests. Thus, the effect of task involvement load on the retention of vocabulary learning needs to be investigated further.

2. The Different Impact of Task Involvement Components

This study attempted to examine the effects of different components of task involvement load on students' vocabulary learning and retention. Laufer and Hulstijn (2001) argued that tasks with the same task involvement load index are conceived as having the same involvement load. Yet, they also suggested that the three components might not have equal impacts. With respect to this suggestion, the second research question was addressed by comparing the mean scores for Tasks 3 and 4.

Before conducting the study, the assumption was that, if Task 3 (sentence writing) and Task 4 (gap-filling with a dictionary), with the same involvement load (3) but different distribution of components, had the same effect on vocabulary learning, there would be no significant differences between the two groups' mean scores. In such a case, the components search and evaluation that distinguished Tasks 3 and 4 would be concluded to have the same relative impact on vocabulary learning.

The results of the immediate posttests confirmed the above assumption, as there was no significant difference in mean scores between the group performing Task 3 (moderate need, no search, and strong evaluation) and that performing Task 4 (moderate need, moderate search, and moderate evaluation). Although these two tasks comprised different components, they had the same involvement load index (3) and resulted in the same effect on initial vocabulary learning. In other words, the components search and evaluation, which distinguished between the two tasks, may have the same relative impact on initial vocabulary learning.

This result is partially similar to those of Youjin Kim (2008), whose experiment 2 investigated the effects of two writing tasks with identical involvement loads (3), namely sentence writing and composition. She claimed that the equivalent involvement load of the two tasks resulted in their equal impact on vocabulary learning. As with Youjin Kim's (2008) claim, for initial vocabulary learning, the sentence writing and gap-filling tasks in the present study, both with an involvement load of 3, had equal impact regardless of their different distribution of components.

In contrast, the delayed posttest results partially supported the above assumption. The findings showed a significant difference between mean scores of Task 3 and Task 4 groups for the delayed posttest of receptive knowledge. This implies that, despite the same involvement load index, the sentence writing task had better vocabulary retention outcome than did the gap-filling task. This finding also suggests that the evaluation component included in the sentence writing task may have a stronger effect on vocabulary retention than does the search component.

This result provides evidence to address the issue suggested by Laufer and

Hulstijn (2001) and Youjin Kim (2008). Although Laufer and Hulstijn (2001) assigned the same relative impact to each component, they also suggested that the three components might not have the same effect on vocabulary learning. In addition Youjin Kim (2008) indicated that tasks with the same involvement index but different component distribution may not be equally effective in vocabulary learning. Based on the current findings regarding research question 2, it is suggested that tasks with the same involvement load but different distribution of components may affect the retention of vocabulary differently. Moreover, evaluation may have a greater impact on involvement load and thus a stronger effect on vocabulary retention than search.

According to Laufer and Hulstijn (2001), information gained through deeper processing is retained longer in long-term memory than is information gained by shallower processing. This implies that a task requiring evaluation may be more effective or longer vocabulary retention in terms of providing the opportunity for deeper processing.

There remains a lack of empirical evidence regarding this issue. In order to reveal the different impact on involvement load of each component, and to determine a reasonable level of involvement for vocabulary learning specifically, further research is required.

VI. CONCLUSION

Learning vocabulary is a fundamental part of learning English. It is not only related to the particular word, but also involves language skills, ideas, and text (Nation, 2011). To encourage vocabulary learning, many researchers, teachers, and curriculum developers have tried to find effective ways of learning.

Laufer and Hulstijn (2001) proposed the involvement load hypothesis, focusing on how the conditions of a task affect incidental learning for vocabulary. The findings of this study partially support the involvement load hypothesis, in that the greater the involvement load, the more effective the vocabulary learning. In other words, a task with higher involvement required deeper processing, allowing learners to benefit from incidental learning in acquiring vocabulary knowledge. According to the present results, the sentence writing task, in which the involvement load was the greatest (3), was the most effective task for initial vocabulary learning. Learners who had completed the reading comprehension task achieved lower scores than did those who had done the other tasks. This result indicates that the generally used task, reading comprehension, may need to be reconsidered in developing learners'

vocabulary knowledge.

With respect to the degree of load, Laufer and Hulstijn (2001) allocated the same level of involvement to three different components (need, search, and evaluation). This study, however, found that two tasks (sentence writing and gap-filling with a dictionary) with the same involvement load index but different components had the same effect on initial vocabulary learning but significantly different effects on the retention of vocabulary. According to these results, the sentence writing task had a more beneficial effect on the retention of words. This means that evaluation, involved in Task 3, had a stronger impact on vocabulary retention than did search, involved in Task 4. In other words, learners engaged in deeper processing for the evaluation component, and thereby achieved better retention of words. With regard to this issue, further research should be carried out to determine a reasonable level of involvement load, as well as the relative weight of each component.

In conclusion, this study has several pedagogical implications. First, teachers need to consider involvement load in the design of tasks targeting learners' vocabulary acquisition. They should be aware that a task with higher involvement, such as a sentence writing task, may be more beneficial, particularly for initial learning. Secondly, for initial learning, tasks with the same involvement load index may be equally beneficial. Thus, as Youjin Kim (2008) implied, learners should not be limited to the use of a certain task. Specifically, for low-proficiency learners, a gap-filling task with a dictionary may be useful for initial vocabulary learning, rather than a writing task, which might be too difficult for them. Thirdly, in order to improve the retention of new vocabulary, writing tasks involving the component of evaluation ought to be considered. Learners should be provided many opportunities to perform writing tasks in order to retain vocabulary knowledge for a longer period.

There are several limitations to the current study. First, the participants' English proficiency was advanced, and so they might have been familiar with the types of vocabulary tests used for the posttests. Thus, they might have used certain strategies to memorize words for the tests. Second, to examine the different degrees of involvement load, a more elaborate task design is needed. In addition, the effect of the mode should be considered in explaining the different effects of tasks in the future studies.

REFERENCES

- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52(2), 323-363.
- Cairns, H. S., Cowart, W., & Jablon, A. D. (1981). Effects of prior context upon the integration of lexical information during sentence processing. *Journal of Verbal Learning and Verbal Behavior*, 20, 445-453.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory research. *Journal of Experimental Psychology: General*, 104(3), 268-294.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, 44(3), 449-491.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40, 273-293.
- Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21(2), 225-241.
- Huh, Jinhee, & Im, Byungbin. (2008). The effects of vocabulary task-induced involvement on vocabulary learning and reading comprehension. *The Journal of Modern British & American Language & Literature*, 26(3), 231-250.
- Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. 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, 539-558.
- Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, 11, 149-158.
- Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19, 357-377.
- Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, 12(3), 365-386.

- Kim, Hyungsun, & Na, Yoonhee. (2010). Vocabulary learning and task-induced involvement. *Korean Journal of Applied Linguistics*, 26(4), 183-211.
- Kim, Jieyoung. (2015). The effects of task-induced involvement and learners' English proficiency on English vocabulary learning. *English Language Teaching*, 27(3), 125-149.
- Kim, Youjin. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58(2), 285-325.
- Knight, S. (1994). Dictionary: The Tool of Last Result in Foreign Language Reading? A New Perspective. *Modern Language Journal*, 78(3), 285-99.
- Laufer, B. H., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694-716.
- Laufer, B. H., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Lee, Yeonkyoung, & Kim, Jieyoung. (2014). Effects of task-induced involvement on EFL learners' vocabulary learning. *The Journal of Linguistic Science*, 72, 297-318.
- Lupescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, 43(2), 263-279.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519-533.
- Nation, I. S. P. (2011). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Schmitt, N. (2004). *Vocabulary in language teaching* (4th ed.). London: Cambridge University Press.
- Shin, Dongkwang, & Chon, Yuh Vicky. (2011). A corpus-based analysis of curriculum-based elementary and secondary English textbooks. *Multimedia-Assisted Language Learning*, 14(1), 149-175.
- Sung, Hyunmi. (2013). Task-induced involvement load in Korean EFL incidental vocabulary learning. *The Journal of Studies in Language*, 29(2), 269-296.

Applicable Levels: College

Jooeun Hyun
Department of English Education
Chung-Ang University
06974 84 Heukseok-ro, Dongjakgu, Seoul, Korea
Tel: (02) 820-5379
Email: hyerang9021@daum.net

Jieyoung Kim
Department of English Education, 203-407
Chung-Ang University
06974 84 Heukseok-ro, Dongjakgu, Seoul, Korea
Tel: (02) 820-5379
Email: jk221@cau.ac.kr

Received December 21, 2015

Revised January 27, 2016

Accepted February 1, 2016