

The Effects of the Sentence-writing Task on English Vocabulary Learning of Korean High School Students*

Jeeyoung Park
(Gyeonggi International Trade High School)

Park, Jeeyoung. (2018). The Effects of the Sentence-writing Task on English Vocabulary Learning of Korean High School Students. *Foreign Language Education Research*, 23, 65-85.

This study was conducted to investigate the effects of the sentence-writing task on English vocabulary learning of Korean high school students. The effectiveness of the sentence-writing task can be explained based on the Involvement Load Hypothesis (Laufer & Hulstijn, 2001); yet, there has been no consistency in the results of the previous studies on this hypothesis. In the present study, the effects of the sentence-writing task on vocabulary learning were re-examined in comparison with the gap-filling task. In addition, considering that there have been no studies addressing the effects of autobiographical elaboration (relating the meaning of a certain word to one's own experience) on memory, the effects of the autobiographical sentence-writing task were compared to those of the imaginary sentence-writing task. Forty high proficiency and 40 low proficiency learners were randomly assigned either of the sentence-writing or the gap-filling task. The results demonstrated that the sentence-writing task is more effective in vocabulary learning than the gap-filling task, regardless of the learners' proficiency levels. However, no significant difference was found between the effects of the autobiographical sentence-writing task and the imaginary sentence-writing task. Based on the results, the pedagogical implications were discussed in the conclusion chapter.

Key Words: vocabulary learning, the sentence-writing task, the Involvement Load Hypothesis, autobiographical elaboration

I . Introduction

Vocabulary is one of the most important factors in communication in that most information indispensable for understanding and expressing meaning is conveyed through vocabulary. According to Wilkins (1972), "While without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111). Nation (2006) and Schmitt (2008) also emphasized the importance of vocabulary knowledge, stating that 98%

* This article was partially extracted from the author's master's thesis.

of the vocabulary in a text should be already known for comprehension, and 8,000 to 9,000 word-family vocabulary is needed for comprehension of written texts and 6,000 to 7,000 for spoken texts. This amount of vocabulary is not easy for learners to acquire, particularly in the EFL environment, where the exposure to L2 input is limited, compared to the ESL environment.

Nevertheless, the best way to improve vocabulary knowledge is not clear yet because of a wide range of variables (Ellis, 1994). As a result, teachers and materials generally fail to provide clear guidelines about vocabulary learning (Schmitt, 2008), and it is not surprising that many learners often indicate the lack of vocabulary knowledge as one of the most crucial reasons for the difficulty in communicating in English.

In an attempt to provide clear descriptions and guidelines about vocabulary learning, the Involvement Load Hypothesis was proposed by Hulstijn and Laufer (2001). According to this hypothesis, the writing task is very effective in vocabulary learning since it induces deeper processing of the target word, requiring strong mental efforts to process it. The Involvement Load Hypothesis suggests three constructs of the effective task in vocabulary learning: *need* (the necessity to know the meanings of the target word), *search* (making efforts to find meanings of the new word), and *evaluation* (the process of making decisions or judgments on appropriateness and relevance of context for the target word). Among these three constructs, *evaluation* index can provide an explanation of the effectiveness of the writing task; since the writing task induces the strongest *evaluation* by requiring learners to formulate a sentence by putting the target word in the appropriate and relevant context, it increases learning efficiency.

A number of studies have been conducted to test the validity of the Involvement Load Hypothesis, particularly regarding *evaluation* index (Cho & Ma, 2015; Eckerth & Tavakoli, 2012; Hulstijn & Laufer, 2001; Jing & Jianbin, 2009; Keating, 2008; Kim, 2008; Lee, 2006; Kim & Na, 2010; Park & Oh, 2015; Sung, 2013; Yang, 2015); however, they have not provided consistent results. They compared the three tasks inducing different *evaluation* indices: the reading-only task (*evaluation* 0), the gap-filling task (*evaluation* 1), and the writing task (*evaluation* 3); yet, their results were different. Although the writing task and the gap-filling task were confirmed to be more effective than the reading-only task, the superiority of the writing task to the gap-filling task was not confirmed in some studies (Kim & Na, 2010; Lee, 2006; Park & Oh, 2015; Soleimani & Rahmanian, 2015; Sung, 2013). In these studies, the learners who performed the writing task did not show better vocabulary gains, compared to those who performed the gap-filling task. Particularly, Lee's (2006) and Sung's (2013) studies demonstrated that the writing task was less effective than the gap-filling task in low proficiency learners' vocabulary learning, which deviates from the Involvement Load Hypothesis. The inconsistency in the results of the previous studies implies that the effects of the writing task on vocabulary learning

need to be re-examined in comparison with the gap-filling task, which induces lower involvement load.

One of the main reasons for this inconsistency might be attributed to the fact that many previous studies did not include the participants' proficiency level as an independent variable. The proficiency level was not fully explored in many studies, except for a few (e.g., Kim, 2008; Kim & Na, 2010; Lee, 2006; Sung, 2013; Yang, 2015). Considering that the learners' proficiency level can be one of the most significant factors that affect vocabulary gains (Kim & Na, 2010; Lee, 2006; Sung, 2013, Yang, 2015), it should be included as a variable in designing such research.

Moreover, there has been a lack of attention to the semantic aspects of the sentences written by learners. For example, in the field of psychology, autobiographical elaboration was found to enhance the memory of target items, as compared to those processed not in relation to the self (Bower & Gilligan, 1979; Holland & Kensinger, 2010; Macrae, Moran, Heatherton, Banfield, & Kelley, 2004; Maki & McCaul, 1985; Reeder, McCormick, & Esselman, 1987; Rogers et al., 1977). Bower and Gilligan (1979) compared the self-reference task (associating the target words with the self) with the other-reference task (associating the target words with Walter Cronkite, a television newscaster), and this experiment revealed that associating the target item with another person (i.e., Walter Cronkite) was less effective in memory than processing the item in relation to the self. This finding was corroborated by some subsequent studies (Maki & McCaul, 1985; Reeder et al., 1987). Reviewing other studies, Holland and Kensinger (2010) stated that "autobiographical memories for personal episodes are often organized into coherent narratives or stories complete with contextual details" (p. 91). Therefore, it seems that autobiographical elaboration needs to be considered as another independent variable set apart from involvement load in exploring the effects of a task on vocabulary learning. In the field of second language vocabulary acquisition, however, no studies have dealt with this semantic aspect of the sentences.

In this context, the present study attempted to examine the effects of the sentence-writing task on Korean high school students' vocabulary learning. The research questions of the present study are as follows:

1. Is the sentence-writing task more effective than the gap-filling task in English vocabulary learning of Korean high school learners at different proficiency levels?
2. Is the autobiographical sentence-writing task more effective than the imaginary sentence-writing task in English vocabulary learning of Korean high school learners at different proficiency levels?

II . Literature Review

1. Overview of General Concepts of Vocabulary Learning

Many studies based on the communicative approach have advocated extensive reading as the effective way for vocabulary learning (e.g., Coady & Huckin, 1997; Krashen, 1989; Pigada & Schmitt, 2006). For example, Krashen (1989) claimed that “competence in spelling and vocabulary is most efficiently attained by comprehensible input in the form of reading” (p. 440). His argument was supported by Coady and Huckin (1997), who claimed that “a great deal of L2 vocabulary is indeed learned through extensive reading” (p. 235). Nation and Ming-Tzu (1999) also emphasized the importance of extensive reading in vocabulary learning, particularly recommending the use of graded readers (i.e., the books organized according to the level of grammatical complexity and vocabulary). Pigada and Schmitt’s (2006) study demonstrated that one month of extensive reading enhanced 65% of the target words’ spelling, meaning, and grammatical characteristics.

Based on many previous studies, it seems to be reliable to some extent that extensive reading contributes to vocabulary development; yet, there have been some contradictions to the effects of extensive reading as well. In particular, many scholars indicated the shortcomings of extensive reading with regard to its inefficiency and the problem of inaccurate guessing (e.g., Laufer, 2005; Nation, 2001; Read, 2004; Peters, Hulstijn, Sercu, & Lutjeharms, 2009; Pressley, Levin, & McDaniel, 1987). For example, Laufer (2005) pointed out the limitation of extensive reading, stating that learners who comprehend the overall message do not attend to the meanings of individual words; moreover, guessing from context is often impossible unless the learner already knows 98% of the words in the reading passage (Laufer, 2005; Nation, 2006; Schmitt, 2008). Another problem is that learners need to encounter the new words quickly and frequently to avoid forgetting them, which is not always possible (Laufer, 2005; Nation, 2001; Read, 2004). Laufer (2005) stated that “in order for words to be met 10 times in reading, learners would need to read 1-2 graded readers per week. The typical learner simply does not read this much” (p. 341). As well as the low pick-up rate, error-prone guessing skill was also pointed out as the problem of extensive reading (Sökmen, 1997). Some studies have shown that learners seldom guess the correct meanings of the target words (Pressley et al., 1987). Reviewing the previous studies, Peters et al. (2009) stated that learning new words only by extensive reading is a “slow and error-prone process” (p. 114); Schmitt (2008) also contended that vocabulary learning programs need to include both extensive exposure and an explicit, intentional learning.

One of the noteworthy hypotheses to address the limitations of extensive reading is the Involvement Load Hypothesis (Hulstijn & Laufer, 2001). This hypothesis offers some

guidelines for vocabulary teaching and learning by reporting the effectiveness of performing tasks that direct attention to specific lexical items.

2. The Involvement Load Hypothesis

In an attempt to provide a clear guideline for vocabulary learning, the Involvement Load Hypothesis was proposed by Laufer and Hulstijn (2001). The Involvement Load Hypothesis categorized three constructs of the effective task in vocabulary learning: *need*, *search*, and *evaluation*. *Need* refers to the necessity for learners to know the meanings of the target words; *search* means whether the learners make efforts to find meanings of the unknown words; *evaluation* refers to the process of making decisions and judgments on relevance or appropriateness of context for the words. According to this hypothesis, learners can acquire vocabulary effectively when they feel the necessity to learn, make efforts to find the meanings of the unknown words, and decide the appropriateness of the context for the target vocabulary.

Many previous studies on the Involvement Load Hypothesis (e.g., Cho & Ma, 2015; Eckerth & Tavakoli, 2012; Hulstijn & Laufer, 2001; Jing & Jianbin, 2009; Keating, 2008; Kim, 2008; Kim & Na, 2010; Lee, 2006; Park & Oh, 2015; Soleimani & Rahmanian, 2015; Sung, 2013; Yang, 2015) particularly have been focused on *evaluation* index. *Evaluation* refers to making a selective decision on the additional words that will combine with the new word, and has three degrees of prominence: no (0), moderate (1) and strong (2). When learners are asked to choose appropriate contexts regarding the target words among the given options, *evaluation* is moderate, and when the learners are required to formulate or construct new sentences or texts on their own using the target words, *evaluation* is strong. The writing task includes the judgment on “semantic and formal appropriateness of the word and its context” (Laufer & Hultijn, 2001, p. 15); thus, according to the Involvement Load Hypothesis, the writing task is beneficial to vocabulary learning since it induces strong *evaluation*.

The previous empirical studies on the Involvement Load Hypothesis compared three tasks that induce different *evaluation* indices: the reading-only task (*evaluation* 0), the gap-filling task (*evaluation* 1), and the writing task (*evaluation* 2). As a result, the writing task and the gap-filling task were confirmed to be more effective than the reading-only task. Yet, the superiority of the writing task to the gap-filling task was not confirmed yet. While some studies (Cho & Ma, 2015; Eckerth & Tavakoli, 2012; Hulstijn & Laufer, 2001; Jing & Jianbin, 2009; Keating, 2008; Kim, 2008; Yang, 2015) provided evidence for the Involvement Load Hypothesis, other studies (Kim & Na, 2010; Lee, 2006; Park & Oh, 2015; Soleimani & Rahmanian, 2015; Sung, 2013) demonstrated that there was no significant difference between the writing task and the gap-filling task. For example, in

Lee's (2006) study, the lower level students showed better vocabulary learning in the gap-filling task, of which *evaluation* index is lower than that of the writing task. Sung's (2013) research showed almost the same results as Lee's (2006) study in that the low proficiency learners benefitted to a greater extent from the gap-filling task rather than the sentence-writing task in vocabulary learning. Kim and Na's (2010), Park and Oh's (2015), and Soleimani & Rahmanian's (2015) studies demonstrated the results deviating from the Involvement Load Hypothesis as well. In these studies, the difference between the sentence-writing task and the gap-filling task was not significant although the learners who completed the sentence-writing task showed slightly better vocabulary gains than those who performed the gap-filling task.

3. Autobiographical Elaboration

Autobiographical Elaboration has not been addressed in the field of second language acquisition yet although many psychological studies have confirmed it as a significant factor in memory. The Autobiographical Elaboration Hypothesis was suggested by Rogers et al. (1977), extending the notion of elaboration to the realm of the 'self'. In Rogers et al.'s (1977) research, the participants were asked to rate the trait adjectives (e.g., *ambitious, hard-working, orderly*) in four tasks designed for different kinds of encoding: structural encoding (rating the size of letters), phonemic encoding (judging the rhyme of the word), semantic encoding (judging the synonym of the word), and self-reference encoding (judging whether the word describes the participants themselves). The incidental recall of the rated words revealed that the items elaborated in the self-reference task were recalled the best. Based on this experiment, Rogers et al. (1977) concluded that when the participant refers a trait adjective to him or herself, the process yields "a rich and powerful encoding" (p. 684) of the word.

Subsequent studies corroborated Rogers et al.'s (1977) finding (e.g., Bower & Gilligan, 1979; Holland & Kensinger, 2010; Macrae et al., 2004; Maki & McCaul, 1985; Reeder et al., 1987; Sui & Humphreys, 2015). Bower and Gilligan (1979) compared the self-reference task with the other-reference task, which required associating the target words with Walter Cronkite, a television newscaster. This experiment revealed that the task that related the target items to another person was less effective in memory because too little was known about him. In this experiment, the items encoded in relation to the self were better remembered since the participants had more knowledge about themselves than they had about other people. Several subsequent studies also demonstrated that when individuals were asked to elaborate stimuli in regard to themselves, memory was enhanced, as compared to when the items were processed for meaning but not in relation to the self (e.g., Holland & Kensinger, 2010; Macrae et al., 2004; Maki & McCaul, 1985; Reeder et

al., 1987).

Reviewing the previous studies on autobiographical elaboration, Holland and Kensinger (2010) stated that “autobiographical memories for personal episodes are often organized into coherent narratives or stories complete with contextual details” (p. 91). Their argument can be supported by Conway and Pleydell-Pearce’s (2000) *self-memory system*, which explains the specificity of autobiographical knowledge. This model suggests that autobiographical knowledge is very specific since it is arranged hierarchically, from the most general information (e.g., when I was in college) to the most specific information (e.g., the spatial layout of the food on the blanket at a picnic). Since autobiographical knowledge includes the information at the most specific level, it may be better remembered than other types of knowledge. Considering these psychological studies, it seems that autobiographical elaboration needs to be considered as another independent variable set apart from involvement load in evaluating the effects of the sentence-writing task on vocabulary learning.

III. Methodology

1. Participants

This research, conducted in July 2015, involves 280 Korean 11th grade high school students. The participants are enrolled in S high school, located in Gyeonggi province. Most of the students had learned English for 8 years in elementary, middle, and high schools, and there was no student who had stayed for more than 6 months abroad.

The initial number of the participants in total was 280 from 8 intact classes. Among the 8 classes, 4 classes were assigned the gap-filling task and the other 4 classes, the sentence-writing task. Among the sentence-writing classes, 2 classes were selected for the autobiographical sentence-writing task (writing about the self) and the other 2 classes were chosen for the imaginary sentence-writing task (writing about an imaginary person). Among the 280 participants, 16 students did not participate in the delayed post-tests, 9 students reported that they already knew some of the target words, and 25 students reported that they reviewed some target words between the immediate and the delayed post-tests; thus, the data of these students were excluded from the analysis.

Considering that the learners’ proficiency level was found to be one of the most significant factors that affect vocabulary gains (Kim & Na, 2010; Lee, 2006; Sung, 2013; Yang, 2015), the selected participants were divided into two proficiency groups: the high proficiency (HP) group, whose percentile scores of the practice CSAT were within 70-99%, and the low proficiency (LP) group, whose percentile scores ranged from 1 to 30%. In order to enlarge the gap between the two proficiency groups, the data of those whose

percentile scores were within 31-69% was excluded from the analysis.

Among all kinds of groups, the number of the HP learners who participated in the imaginary sentence-writing task was only 10, the smallest number. Thus, to equalize the size of the different proficiency groups in the same task, 10 students were randomly selected from the LP learners in the imaginary sentence-writing group. Subsequently, in order to equalize the size of the imaginary sentence-writing group and the autobiographical sentence-writing group, 20 students were randomly selected from the autobiographical sentence-writing group: 10 HP learners and 10 LP learners. As a result, the total number of the data for the sentence-writing task was 40.

Finally, to equalize the size of the sentence-writing group and the gap-filling group, 40 students were randomly selected from the learners who performed the gap-filling task: 20 HP learners and 20 LP learners. As a result, 80 participants were selected for the analysis. The homogeneity of each task group was confirmed by one-way ANOVAs (all $p > .05$).

TABLE 1
Distribution of the Participants of the Experimental Groups

Proficiency Level	Gap-filling	Sentence-writing		Total
		Autobiographical	Imaginary	
HP	20	10	10	40
LP	20	10	10	40
Whole	40	20	20	80

2. Instruments

1) Target Words

Ten target words were chosen based on three criteria: (1) the assumed unfamiliarity to the participants; (2) the ease of writing sentences by using the words; (3) the ease of translating into Korean. With regard to the second criterion, the words considered to be too hard for high school students to make an appropriate sentence were ruled out, such as *explicit*, *intrinsic*, or *equilibrium*, of which meanings may be beyond their common knowledge. In order to avoid selecting such technical words, most of the target words were chosen from the novels that consist of daily vocabulary such as *The Adventure of Huckleberry Finn* and *Gulliver's Travels*. In terms of the third criterion, the words that were deemed to be difficult to translate into Korean were excluded.

Following the previous studies (Kim, 2008; Kim & Na, 2010; Sung, 2013; Yang, 2015), which used 10 new words, 10 target words were selected in the following way. First, a

total of 30 candidate target words (10 verbs, 10 adjectives, 10 nouns) were chosen by the researcher. After the discussion with the teacher who had taught the participants for one and half years, seven words that the students had already learned were ruled out. Next, with the rest of the words, a pretest was conducted on 20 high proficiency students from another school. As a result, 10 words that none of the students knew were selected as the target words: 3 verbs (*irk, mar, upbraid*), 3 adjectives (*insipid, surly, voluble*), and 4 nouns (*bungle, chum, prowess, squabble*).

2) Tasks

(1) Gap-filling Task

The participants in the gap-filling task group were asked to fill in the blanks in the given sentences. The meaning and the part of speech were presented in a wordlist. In addition to the 10 target words, 2 distracters (*dissent* and *remarkable*) were added to keep the students from guessing the meanings of the last words; if the number of target words and that of blanks were the same, the participants would not make a careful judgment for the last one or two words (Yang, 2015). The total 12 words were presented in an alphabetical order.

(2) Autobiographical Sentence-writing Task

The sentence-writing task was divided into two types: the autobiographical sentence-writing task and the imaginary sentence-writing task. In the autobiographical sentence-writing task, the students were asked to write one or two sentences about their own experiences, including each target word. Some previous studies provided students with an example sentence for each target word since the researchers thought that the low proficiency students would not be competent enough to produce sentences without any guidelines (Lee, 2006; Yang, 2015). Nevertheless, the current study did not provide any example sentences to avoid the possibility that the students would imitate the content of the example sentences rather than creating their own. Instead of providing example sentences, the current study included specific guidelines for assisting the students in brainstorming the content of each sentence. For example, in the case of the word *surly*, the guideline “Write about your own experience where you saw a *surly* person. Please describe how the person behaved in the *surly* way and towards whom.” was provided in Korean.

(3) Imaginary Sentence-writing Task

In the imaginary sentence-writing task, the students were asked to write one or two sentences about the imaginary person *Tom*, including each target word. For example, in the case of the word *surly*, a guideline such as “Write one or two sentences where Tom is

a kind of *surly* person. Please imagine and describe how he behaved in the *surly* way and towards whom.” was presented in Korean. As in the autobiographical sentence-writing task, no example sentences were provided.

3) Assessment

(1) Active Word Learning Test

In the active word learning test, the learners were provided with the list of 10 target words in Korean and were required to write the equivalent English words. As a scoring rubric, the lexical production scoring protocol was selected, which was suggested by Barcroft (2002). It enables a more detailed analysis of the students’ vocabulary knowledge by providing specific scores to the developing partial forms as well as the complete word forms. Given that the post-tests were conducted after only one exposure to the target words, the scoring system that is sensitive to the slight differences in the learners’ initial learning may be more appropriate (Yang, 2015). Scoring was conducted by the researcher of the study and an English teacher who had taught English for 4 years in middle and high school. Pearson’s r was calculated to check the inter-rater reliability, and the attained value was 0.983. Because of the high inter-rater reliability, one of the rater’s scores was randomly selected for data analysis.

(2) Passive Word Learning Test

In the passive word learning test, the participants translated the 10 target words from English to Korean. One point was given for each correct answer and 0.5 point was given to semantically acceptable answers, based on the previous studies (Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Yang, 2015). The words that were correct in meaning but incorrect in the part of speech were given 0.5 point. Confusion and disagreement were solved through discussions between the raters. After each rater finishing scoring, the inter-rater reliability was calculated with Pearson’s correlation, and the attained value was 0.987. Since the inter-rater reliability was very high, one of the raters’ scores was randomly selected and used for the data analysis.

4) Procedure

Two pilot studies were conducted to select an appropriate sentence-writing task for the main study. The participants in the pilot studies were 42 students from H high school in Gyeonggi province. Their percentile scores in the practice CSAT were within 40-60%; thus, their English proficiency was at the intermediate level.

(1) Pilot Study

The participants in the first pilot study were 28 students, who were asked to write one or two sentences with four target words: *bungle*, *chum*, *insipid*, and *irk*. The participants were divided into two groups: the group A and the group B. Fourteen students in the group A wrote sentences without any content guidelines while the other 14 students in the group B were provided with the guidelines such as “Write about your own experience where you made a *bungle*. Please describe when and what *bungle* you made.” To ensure the students’ comprehensibility, the guidelines were provided in Korean. As a result of comparing the length of the sentences written by both groups through a *T*-test, a significant difference was found ($T=4.233$, $p<.05$). The learners in the group B, who were provided with the content guidelines, created longer context than those in the group A. This finding provided the researcher the rationale to include the content guidelines for the main study participants.

The second pilot study was conducted to determine whether to provide example sentences for the sentence-writing group in the main study. The group B mentioned above was compared with the new group C consisting of 14 students. As described above, the group B in the first pilot study wrote sentences with content guidelines, but no example sentences were provided for them. In contrast, the group C was provided with both the content guidelines and the example sentences. Thus, the only difference between the group B and C was whether they were provided with the example sentences or not. The comparison between the sentences of the two groups revealed that the students in the group C copied many parts of the example sentences. For example, the example sentence for the word *irk* was “My brother *irked* me yesterday. He made a noise when I was studying.”, and 10 students among 14 in the group C included the phrase *make a noise* in their sentences. Based on this result, it was concluded that providing example sentences may not be helpful in encouraging students to write their own stories since they tend to imitate the given sentences. Therefore, no example sentences were provided for the learners in the main study.

(2) Main Study

The procedure of the main study was divided into two sessions. In the first session, 280 learners completed the assigned tasks and took the immediate post-tests. The researcher was present during the entire process, and the directions on the tasks were written on the task sheets in Korean. The students were encouraged not to concentrate too much on grammatical accuracy. The time spent on the tasks was not controlled, on the basis of Hulstijn and Laufer’s (2001) claim that time is “an inherent property of a task” (p. 549). The researcher wrote the starting time on the blackboard, and the participants were asked to write the finishing time right after they completed the task. The time spent on each task was calculated by the researcher, and the average task completion time of the

HP learners was 5.20 minutes, 18.93 minutes, and 20.90 minutes for the gap-filling task, the autobiographical sentence-writing task, and the imaginary sentence-writing task, respectively. For the LP learners, the gap-filling task took 6.87 minutes, the autobiographical sentence-writing task, 19.33 minutes, and the imaginary sentence-writing task, 20.33 minutes on average.

To prevent rote learning, the learners were not preannounced about the upcoming immediate post-tests. After all class members completed the tasks, the task sheets were collected and the immediate post-tests were conducted. The immediate post-tests consisted of two parts: the active test (writing the target words corresponding to the Korean meanings) and the passive test (writing the Korean translation of the target words). To prevent the influences of the previous test on the next test, the active test was provided ahead of the passive test (Webb, 2005). Students' answer sheets were collected after three minutes, when all the students completed the task.

The learners were provided with the delayed tests one week after the first session, following Hulstijn and Laufer's (2001) method. The order of the items on the sheets was changed from that of the immediate tests to prevent the memory of the order of each item from influencing the results. At the end of the delayed passive test, the learners were asked to check the words that they had reviewed between the first session and the second session.

5) Data Analysis

For the first research question regarding the effects of the sentence-writing task and the gap-filling task on vocabulary learning, a series of two-way ANOVAs were conducted with the task type (the sentence-writing task and the gap-filling task) and the proficiency level (HP and LP) as the independent variables and the four post-test scores (the immediate active, the immediate passive, the delayed active, and the delayed passive tests) as the dependent variables. The two-way ANOVAs were conducted using *SPSS 20*, and the significant level was set at 0.05.

Likewise, to answer the second research question, a set of two-way ANOVAs were conducted with the task type (the autobiographical sentence-writing task and the imaginary sentence-writing task) and the proficiency level (HP and LP) as the independent variables and the four post-test scores (the immediate active, the immediate passive, the delayed active, and the delayed passive tests) as the dependent variables.

IV. Results and Discussion

1. The Effects of Involvement Load on Vocabulary Learning

In order to examine whether the sentence-writing task is more effective than the task inducing lower involvement load, that is, the gap-filling task in vocabulary learning, the participants (40 HP learners and 40 LP learners) performed one of the two tasks and took the immediate and the delayed post-tests. Table 2 demonstrates the means and the standard deviations of the post-test scores of the different proficiency groups (HP, LP, and the whole group combining the two proficiency groups).

TABLE 2
Descriptive Statistics of the Post-test Scores

Proficiency	Task	Immediate Test				Delayed Test				N
		Active		Passive		Active		Passive		
		M	SD	M	SD	M	SD	M	SD	
HP	Gap-filling	2.76	1.76	3.93	2.17	1.14	1.10	2.33	1.66	20
	Sentence-writing	5.09	2.23	7.53	2.54	2.45	1.72	4.88	2.99	20
LP	Gap-filling	0.76	1.12	1.48	1.25	0.01	0.06	0.48	0.55	20
	Sentence-writing	3.74	2.68	5.38	3.07	1.73	1.68	3.03	2.18	20
Whole	Gap-filling	1.76	1.77	2.70	2.14	0.58	0.96	1.40	1.54	40
	Sentence-writing	4.41	2.53	6.45	2.99	2.09	1.72	3.95	2.75	40

Note. The maximum score for each test is 10.

The result shows that the sentence-writing task yields better vocabulary gains than the gap-filling task in both initial learning and long-term retention regardless of the learners' proficiency levels. It also indicates that the sentence-writing task is more effective than the gap-filling task in both active word learning and passive word learning. In the three proficiency groups, the mean scores of the sentence-writing group are higher than those of the gap-filling group in every case.

In order to check whether there is a significant difference between the two tasks and whether there is an interaction between the task type and the proficiency level, two-way ANOVAs were conducted. The effects of the task on the post-test scores were found to be significant in every post-test (all $p < .05$), and the proficiency level was also found to have

a significant effect on all post-test scores (all $p < .05$). As for the interaction effect between the task and the proficiency level, no significant interaction was found in every post-test (all $p > .05$). This indicates that the task with higher involvement load has a significant effect on vocabulary learning regardless of the learners' proficiency levels.

This result is different from those of many previous studies conducted in the EFL context (e.g., Kim & Na, 2010; Lee, 2006; Park & Oh, 2015; Soleimani & Rahmanian, 2015; Sung, 2013). In those studies, the difference between the sentence-writing task and the gap-filling task was not found to be significant. In particular, Lee's (2006) and Sung's (2013) studies demonstrated that the gap-filling task was more effective than the sentence-writing task in the LP learners' vocabulary learning.

In terms of Sung's (2013) study, the sentence-writing task did not require the learners to create their own sentences but just to order 10 target words to make sentences equivalent to Korean sentences. In other words, the content of the sentences was already provided by the researcher, and the task provided for the learners was translating the Korean sentences into English by unscrambling the given English words. Although the researcher may have attempted to provide the learners with some scaffolding with regard to the content and vocabulary, the task may have been still difficult for the LP learners, who did not have sufficient knowledge of the English sentence structure. This problem may have led to the result that the LP learners of the sentence-writing group showed the lower rate of vocabulary learning than those who performed the gap-filling task.

When it comes to Lee's (2006) study, some example sentences were provided to assist the participants in writing English sentences; yet, there is a possibility that the students, particularly the LP learners, copied some words or phrases in the example sentences. In the pilot study of the current research, the low proficiency students showed a tendency to copy the example sentences rather than creating their own. Based on this result, it can be concluded that providing example sentences may not be conducive to encouraging students to write their own stories with the target words.

In other previous studies that showed the results deviating from the Involvement Load Hypothesis (e.g., Kim & Na, 2010; Park & Oh, 2015; Soleimani & Rahmanian, 2015), no guideline was provided for assistance; thus, the learners' sentences may have been relatively short. In fact, the pilot study of the current research demonstrated that the students tended to write very short sentences when no content guidelines were provided. Most students who were not provided with any content guidelines wrote short and simple sentences such as "I have a *bungle*," or "She is my best *chum*," while those who were presented with the guidelines completed much longer sentences such as "I was *bungle* to speak carelessly yesterday," or "I love Ye-in because I see a *chum* long time," although their grammar was not perfect. The pilot study indicated that learners can create longer context when they are provided with content guidelines. Based on this pilot study, content

guidelines were provided for the learners in the main study, which may have enabled them to formulate longer context. Considering that many Korean learners do not have a lot of opportunities to write sentences in English with their own ideas, providing them with content guidelines in the sentence-writing task may be conducive to creating longer context, enhancing their vocabulary learning gains.

2. The Effects of Autobiographical Elaboration on Vocabulary Learning

The effects of autobiographical elaboration on vocabulary learning are presented in this section. The analysis was conducted in three proficiency groups: HP, LP, and the whole group. To investigate the effects of autobiographical elaboration on initial vocabulary learning and its long-term retention, 20 learners in the autobiographical sentence-writing group and 20 learners in the imaginary sentence-writing group were compared in terms of their post-test scores. Table 3 displays the means and the standard deviations of the post-test scores of the learners.

TABLE 3
Descriptive Statistics of the Post-test Scores

Proficiency	Task	Immediate Test				Delayed Test				N
		Active		Passive		Active		Passive		
		M	SD	M	SD	M	SD	M	SD	
HP	Autobiographical	4.28	1.73	6.90	2.88	2.70	1.63	4.05	2.41	10
	Imaginary	5.90	2.46	8.15	2.11	2.20	1.87	5.70	3.40	10
LP	Autobiographical	3.88	2.42	6.00	3.12	1.75	2.00	2.95	2.29	10
	Imaginary	3.60	3.04	4.75	3.05	1.70	1.40	3.10	2.18	10
Whole	Autobiographical	4.08	2.06	6.45	2.96	2.23	1.84	3.50	2.36	20
	Imaginary	4.75	2.94	6.45	3.09	1.95	1.63	4.40	3.08	20

Note. The maximum score for each test is 10.

This descriptive result deviates from the Autobiographical Elaboration Hypothesis in that the scores of the imaginary sentence-writing group are higher than those of the autobiographical sentence-writing group in many post-tests. For example, among the HP learners, the students in the imaginary sentence-writing group gained better scores than those in the autobiographical sentence-writing group in most tests except for one, the

delayed active test. Among the LP learners, those of the autobiographical sentence-writing group performed better than those of the imaginary sentence-writing group in the immediate active, the immediate passive, and the delayed active tests. Yet, an opposite pattern appeared in the delayed passive test; that is, the imaginary sentence-writing group showed better performance than the autobiographical sentence-writing group in the delayed passive test. In terms of the whole group, the imaginary sentence-writing group showed better performance than the autobiographical sentence-writing group in the immediate active and the delayed passive tests. These results are different from the hypothesis that the task inducing autobiographical elaboration is more effective than other types of elaboration in memory.

In order to check whether the difference between the tasks is statistically significant and whether there is an interaction between the task and the proficiency level, a set of two-way ANOVAs were conducted, and no significant difference was found between the effects of the two types of sentence-writing tasks (all $p > .05$). This result deviates from the previous studies which demonstrated that autobiographical elaboration enhanced the memory of target items (e.g., Holland & Kensinger, 2010; Kensinger, 2004; Macrae et al., 2004; Reeder et al., 1987; Rogers et al., 1977).

In terms of the learners' proficiency level, its effects on the immediate passive and the delayed passive tests were found to be significant ($p < .05$) while its effects on the other two tests, namely, the immediate active and the delayed active tests were not significant ($p > .05$). This indicates that the students' proficiency level assessed by the practice CSAT does not significantly affect their active word learning.

As for the interaction between the task and the proficiency level, it was not found to be significant (all $p > .05$). This result implies that the effects of autobiographical elaboration on vocabulary learning are not influenced by the learners' proficiency level. Thus, it can be concluded that autobiographical elaboration does not affect vocabulary learning regardless of the learners' proficiency levels. A possible explanation for this can be provided with regard to the resemblance in the content of the sentences. For example, in terms of the target word *bungle*, the mistake to break something (e.g., a window, a cup, a dish, etc.) commonly appeared in the sentences regardless of autobiographical elaboration. Likewise, regarding the word *insipid*, most students in the two sentence-writing groups wrote that studying a certain subject (e.g., literature, mathematics, English, etc.) is *insipid*. This similarity in content may be attributed to the learners' tendency to draw on their own past experiences in performing both types of tasks. This implies that the boundary between the autobiographical sentence-writing task and the imaginary sentence-writing task is not very clear. In the case of the previous studies, many of them compared the autobiographical elaboration task with another task relating the target word to other real people such as the newscaster Walter Cronkite (Bower & Gilligan, 1979),

President Ronald Reagan (Maki & McCaul, 1985), and Princess Diana (Reeder et al., 1987). Since these people actually existed, it was not possible for the participants to make up the stories, based on their own personal experiences; thus, the boundary between the different types of writing tasks may have been clearer than in the current study.

There is another possibility that the results of the current study may be attributed to the fact that elaboration was restricted to only the sentence level. For example, in Reeder et al.'s (1987) study, the experimental group was asked to read a prose consisting of approximately 100 words, thinking whether the story was related to the participants themselves. In contrast, the comparison group was instructed to think whether the story was related to Princess Diana. After they read the prose, they were asked to retell the whole prose that they had read, and it was found that autobiographical elaboration was more effective than the other type of elaboration (i.e., relating the story to Princess Diana) in recalling the story. Unlike Reeder et al.'s (1987) study, the current study limited elaboration to the sentence level, rather than the discourse level; thus, the effects of autobiographical elaboration may not have appeared clearly. In addition, while Reeder et al.'s (1987) study required the participants to recall the whole story that they had read, the present study required them to recall some specific words in the sentences. If the participants of the current study had been asked to recall the whole sentences, the results might have been different.

In addition, the similarity of the two types of sentence-writing in length and concreteness may have influenced the results as well. Holland and Kensinger (2010) claimed that "autobiographical memories for personal episodes are often organized into coherent narratives or stories complete with contextual details" (p. 91); yet, in the present study, the specific content guidelines provided in both writing tasks seem to have made the length and the concreteness of the two types of sentences similar to each other. If no content guidelines had been provided, the autobiographical sentence-writing task might have yielded longer and more detailed sentences than the imaginary sentence-writing task.

V. Conclusion

The research for the first question revealed that the sentence-writing task, which induces higher involvement load, yields better vocabulary learning than the gap-filling task, regardless of learners' proficiency levels. The learners who completed the sentence-writing tasks (the autobiographical sentence-writing task and the imaginary sentence-writing task) showed better active and passive word learning both in the immediate and the delayed tests, than those who performed the gap-filling task. The task type was found not to interact with the learners' proficiency level; that is, both the high and the low proficiency learners benefitted from the sentence-writing task in vocabulary learning. This

result deviates from many previous studies conducted in the EFL context which demonstrated that there was no significant difference between the sentence-writing task and the gap-filling task in terms of their effects on vocabulary learning (e.g., Kim & Na, 2010; Park & Oh, 2015; Soleimani & Rahmani, 2015). In the current study, the sentence-writing group was provided with the content guidelines, which may have been conducive to creating longer sentences. Many Korean EFL learners may not be used to expressing their ideas or thoughts in English since most of the writing tasks provided in school are just writing a single word or short phrases. The pilot study of the current research confirmed the learners' tendency to write very short and simple sentences when no content guidelines are provided for them. Based on this result, some content guidelines were provided for the sentence-writing groups in the main study, which may have assisted the learners in creating long sentences. This finding provides an implication that providing specific guidelines for content might be necessary to improve the effects of the sentence-writing task on vocabulary learning.

As for the effects of autobiographical elaboration, no significant difference was found between the autobiographical sentence-writing task and the imaginary sentence-writing task. This implies that whether to write about one's own experiences or about imagined stories does not affect vocabulary learning. This finding deviates from the previous studies which demonstrated that autobiographical elaboration enhanced memory (e.g., Bower & Gilligan, 1979; Holland & Kensinger, 2010; Macrae et al., 2004; Maki & McCaul, 1985; Reeder et al., 1987; Rogers et al., 1977). When looking into the sentences written by the learners, some common themes often appeared in both the autobiographical and the imaginary stories; thus, it seems that learners tend to draw on their own past experiences no matter what types of sentence-writing tasks are assigned to them. This implies that the boundary between the autobiographical elaboration task and the imaginary elaboration task is not very clear. This finding may provide the ground for utilizing both types of writing tasks in vocabulary teaching and learning. If learners have difficulty in brainstorming their own experiences related to the target word, it would be better to allow them to write sentences with imagined stories. The imaginary sentence-writing task would yield almost the same results as the autobiographical sentence-writing task on vocabulary learning.

Given the scope of the current study, there are some suggestions for future research. First, the post-tests utilized in the current study were limited to the spelling and Korean translation of the target words. This word knowledge might be too limited and simplified, considering how complex and deep vocabulary knowledge is. Future research would benefit from examining a wider range of vocabulary knowledge such as grammatical functions, collocations, and register. Another limitation of the current research is that the study was conducted only with the students from one specific high school. Thus, there is

a limitation to generalize this finding to other school students. To improve the reliability of the research, it is recommended that future research be conducted with the students from other high schools. In addition, it is necessary to conduct a similar study with middle school students as well as high school students.

REFERENCES

- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*(2), 323-363.
- Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality, 13*(4), 420-432.
- Cho, Y. A., & Ma, J. H. (2015). Task effectiveness and word learning in a second language acquisition. *English Literature Research, 41*(1), 313-336.
- Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press.
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*(2), 261.
- Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research, 16*(2), 227-252.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of Life Reviews, 7*(1), 88-131.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.
- Jing, L., & Jianbin, H. (2009). An empirical study of the Involvement Load Hypothesis in incidental vocabulary acquisition in EFL listening. *Polyglossia, 16*, 1-11.
- 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, H. S., & Na, Y. H. (2010). Vocabulary learning and task-induced involvement. *Applied Linguistics, 26*(4), 183-211.
- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 285-325.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440-464.

- Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla Yearbook*, 5(1), 223-250.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Lee, Y. H. (2006). *The impact of task-induced involvement load on EFL vocabulary learning*. Unpublished master's thesis. Seoul National University.
- Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex*, 14(6), 647-654.
- Maki, R. H., & McCaul, K. D. (1985). The effects of self-reference versus other reference on the recall of traits and nouns. *Bulletin of the Psychonomic Society*, 23(3), 169-172.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, I. S. P., & Ming-Tzu, K. W. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12(2), 355-380.
- Park, J. Y., & Oh, J. I. (2015). Effects of task-induced vocabulary learning involvement on vocabulary learning by Korean university students. *English Language Teaching*, 27(3), 175-191.
- Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, 59(1), 113-151.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Pressley, M., Levin, J. R., & McDaniel, M. A. (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 107-127). Hillsdale, NJ: Lawrence Erlbaum.
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146-161.
- Reeder, G. D., McCormick, C. B., & Esselman, E. D. (1987). Self-referent processing and recall of prose. *Journal of Educational Psychology*, 79(3), 243.
- Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 35(9), 677.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.

- Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp.237-258). Cambridge: Cambridge University Press.
- Soleimani, H. & Rahmanian, M. (2015). Vocabulary acquisition and task effectiveness in Involvement Load Hypothesis: A case in Iran. *International Journal of Applied Linguistics and English Literature*, 4(5), 198-205.
- Sui, J., & Humphreys, G. W. (2015). The integrative self: How self-reference integrates perception and memory. *Trends in cognitive sciences*, 19(12), 719-728.
- Sung, H. M. (2013). Task-induced involvement load in Korean EFL incidental vocabulary learning. *Language Research*, 29(2), 269-296.
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27, 33–52.
- Wilkins, D. A. (1972). *Linguistics in language teaching*. London: E. Arnold, 1973.
- Yang, S. E. (2015). *The effects of the involvement load of tasks on English vocabulary learning of Korean high school students*. Unpublished masters' thesis. Seoul National University.

Park, Jeeyoung
English teacher
Gyeonggi International Trade High School
141 Seokcheon-ro, Bucheon, Gyeonggi
Tel: (032) 320-6211
Email: jycjn1003@korea.kr

Received on 2018.10.31.

Reviewed on 2018.11.19.

Revised version received on 2018.12.07.

Accepted on 2018.12.08.