교육학석사학위논문

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

문장 쓰기 과제가 한국 고등학생들의 영어 어휘 학습에 미치는 영향

2016 년 2 월

서울대학교 대학원

외국어교육과 영어교육전공

박 지 영

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

by

JEE-YOUNG PARK

A Thesis Submitted to
the Department of Foreign Language Education
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Education

At the
Graduate School of Seoul National University

February 2016

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

문장 쓰기 과제가 한국 고등학생들의 영어 어휘 학습에 미치는 영향

지도교수 오 선 영

이 논문을 교육학 석사 학위논문으로 제출함
2015년 12월

서울대학교 대학원
외국어교육과 영어교육전공
박 지 영

박지영의 석사학위논문을 인준함
2016년 1월

위 원 장 _____

부위원장 _____

위 원 _____

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

APPROVED BY THESIS COMMITTEE:

_____

JIN-WAN KIM, COMMITTEE CHAIR

_____

YOUNGSOON SO

_____

SUN-YOUNG OH

ABSTRACT

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) in that the tasks inducing higher involvement load yield better results in vocabulary learning. A number of studies have been conducted to confirm the Involvement Load Hypothesis; yet, there has been no consistency in their results. The inconsistency mainly appeared in the comparison between the sentence-writing task and the gap-filling task. Some studies have shown the results that there was no significant difference between the effects of the sentence-writing task inducing higher involvement load and those of the gap-filling task inducing lower involvement load. Thus, 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, since there have been no studies that addressed the effects of autobiographical elaboration (relating the meaning of a certain word to one's own experience) on vocabulary learning, the current study compared the effects of the autobiographical sentence-writing task and those of the imaginary sentence-writing task. Additionally, considering that there has been a lack of attention to the differences between the sentences

written by the learners who achieved higher vocabulary retention and the sentences written by those who showed less vocabulary gains, the current study compared the quality of the sentences written by the learners whose vocabulary test scores were different.

In the present study, 40 high proficiency learners and 40 low proficiency learners in one high school located in Gyeonggi province were selected as the participants and randomly assigned either of the sentence-writing task or the gap-filling task. As for the sentence-writing task, the task was divided into the autobiographical sentence-writing task and the imaginary sentence-writing task. In order to assist the students in brainstorming the content of the sentences, some short guidelines were provided in Korean. For example, in the case of the word *surly*, a guideline such as "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.

A set of two-way ANOVAs conducted between the sentence-writing group and the gap-filling group demonstrated the results consistent with the Involvement Load Hypothesis. That is, the sentence-writing task was found to be more effective in vocabulary learning than the gap-filling task, regardless of the learners' proficiency levels. This result seems to be attributed to the content guidelines provided for the sentence-writing group, which may have enabled the

learners to extend the length of their sentences. However, no significant difference was found between the effects of the autobiographical sentence-writing task and those of the imaginary sentence-writing task. This implies that whether the sentence is written based on learners' past experience does not affect vocabulary learning. Lastly, the sentences written by the high and the low post-test score groups were analyzed using *Coh-Metrix 3.0*. and compared through *T*-tests. The *T*-test results revealed significant differences between the two score groups in the length of context, lexical diversity, and the frequencies of adverbs, causal connectives, and negation markers. This finding sheds some light on the possibility that these properties of the sentences may affect vocabulary learning. 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

Student Number: 2012-21372

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1
INTRODUCTION

This chapter introduces the current research in terms of its motivation and organization. The purpose for the study is discussed in Section 1.1, followed by the research questions in Section 1.2. The organization of the thesis is outlined in Section 1.3.

1.1 The Purpose for the Study

Vocabulary is one of the most important factors in communication in that most information indispensible for understanding and expressing meaning is conveyed though vocabulary. According to Wilkins (1972), "While without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111). Sternberg (1987) also stated that "one's level of vocabulary is highly predictive, if not determinative, of one's level of reading comprehension" (p. 90). Likewise, Nation (2006) and Schmitt (2008) emphasized the importance of vocabulary knowledge, stating that 98% 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; de Groot, 2006). 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, 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; Knight, 1994; Lee, 2006; Prince, 1996; Pulio, 2003; Sung, 2013, Swanborn & De Glopper, 2002; Van Daalen-Kapteijns et al., 2001; 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; Brown, Keenan, & Potts, 1986; Holland & Kensinger, 2010; Macrae et al., 2004; Maki & McCaul, 1985; Reeder et al., 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 (Brown, Keenan, & Potts, 1986; 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.

Furthermore, there has been no analysis on the sentences themselves written by learners yet. Although there may be some qualitative differences between the sentences written by the learners who perform better in vocabulary retention and the sentences written by those who show less vocabulary gains, no previous studies have conducted this type of sentence analysis.

In this context, the present study attempted to examine the effects of the sentence-writing task on Korean high school students' vocabulary learning. First, with regard to involvement load, the effects of the sentence-writing task on vocabulary learning were compared with those of the gap-filling task, which induces lower involvement load than the sentence-writing task. Second, regarding autobiographical elaboration, the effects of the autobiographical sentence-writing

task were compared with those of the imaginary sentence-writing task. Lastly, an analysis was conducted with regard to the qualitative aspects of the sentences written by learners.

1.2 Research Questions

As mentioned above, the purpose for the current study is to investigate the effects of the sentence-writing task on vocabulary learning from the following perspectives. First, the current study compares the sentence-writing task with the gap-filling task, in terms of their effects on vocabulary learning. Second, given that the effects of the sentence-writing task can vary according to whether it induces learners' autobiographical elaboration or not, the current study compares the autobiographical sentence-writing task and the imaginary sentence-writing task in terms of their effects on vocabulary learning. As for these two research questions, the learners' proficiency level was considered as another independent variable in designing the experiment. Third, considering that there might be other variables that modify the effects of the sentence-writing task, diverse characteristics of the students' sentences were closely examined. Thus, 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?

3. Is there any difference in the sentences written by the learners who achieved higher versus lower vocabulary retention?

1.3 Organization of the Thesis

This thesis consists of five chapters. Chapter 1 introduces the current research in terms of its motivation and organization. Chapter 2 explains the general concepts of vocabulary learning, the Involvement Load Hypothesis and its theoretical background, and autobiographical elaboration. Chapter 3 presents the methodology in terms of the participants, instruments, assessment, procedure, and data analysis. Chapter 4 describes the results and discusses the findings. Chapter 5 provides the summary of the findings, some implications and limitations of the current study, and the suggestions for further studies.

CHAPTER 2
LITERATURE REVIEW

This chapter presents the theoretical backgrounds of vocabulary learning and two hypotheses that can explain the effects of the sentence-writing task on vocabulary learning: the Involvement Load Hypothesis and the Autobiographical Elaboration Hypothesis. First, Section 2.1 provides an overview of general concepts of vocabulary learning. Next, Section 2.2 discusses the Involvement Load Hypothesis and its theoretical background, the Depth of Processing Hypothesis. Lastly, Section 2.3 discusses autobiographical elaboration and some previous empirical research on it.

2.1 Overview of General Concepts of Vocabulary Learning

In the grammar-translation approach of the 1930s, word lists, definitions, and flash cards were the main elements of vocabulary teaching and learning. As a reaction to the reading approach, the audio-lingual approach emerged, which was based on the assumption that language is habit formation, holding the dominant position from the 1940s through the 1960s. Since 1960s, the teaching of vocabulary was limited because of the primary focus on teaching grammar and

pronunciation. In 1970s, the goal of language teaching was focused on the ability to communicate in the target language, focusing more on fluency rather than accuracy. The syllabus was designed centering on notions and functions, and in this syllabus format, grammar and vocabulary were seen as subsidiary aspects. Furthermore, as the role of teachers was perceived as facilitators and guiders, they were not likely to take directive steps to turn students' focus to lexical form (Brown, 2007). In this method, the implicit and incidental learning of vocabulary was recommended (Sökmen, 1997). L2 students were taught to find clues in context, use monolingual dictionaries, and avoid defining words or glossing texts in their native language. They were instructed to infer word meaning from context, which was regarded as the best vocabulary skill. As a result, the emphasis on extensive reading emerged.

Many studies based on the communicative approach have advocated extensive reading as the effective way for vocabulary learning (e.g., Coady, 1997; Day & Swan, 1998; Elley & Mangubhai, 1981; Grabe & Stoller, 1997; Hafiz & Tudor, 1989; Krashen, 1989; Nagy, Herman, & Anderson, 1985; Nation & Wang, 1999; Pellicer-Sánchez & Schimitt, 2010; 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 (1997), who claimed that "a great deal of L2

vocabulary is indeed learned through extensive reading" (p. 235). Nation and Wang (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., Kelly, 1990; Laufer, 2005; Nation, 2001; Read, 2004; Peters et al., 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; Kelly, 1990). 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), suggested based on the Depth of Processing Hypothesis (Craik & Lockhart, 1972) in the field of psychology. The Involvement Load Hypothesis offers some guidelines for vocabulary teaching and learning by reporting the effectiveness of performing tasks that direct attention to specific lexical items.

2.2 The Involvement Load Hypothesis and Its Theoretical Background

The Involvement Load Hypothesis in the SLA field was derived from the

Depth of Processing Hypothesis in the field of psychology. Section 2.2.1 discusses the Depth of Processing Hypothesis as the theoretical background for the Involvement Load Hypothesis. Next, Section 2.2.2 provides more detailed explanations of the Involvement Load Hypothesis and some empirical research on it.

2.2.1 The Depth of Processing Hypothesis

Craik and Lockhart (1972) first proposed the Depth of Processing Hypothesis, suggesting that the retention in the long-term memory depends on how deep the new information is processed during the learning. According to this hypothesis, when the input is initially processed at the deep level, strong memory traces are built, and the new information is retained longer. This hypothesis was proposed based on several previous findings that encoding an item only at a physical and sensory level without any intention to learn does not facilitate memory performance (Moray, 1959; Norman, 1969). Based on these studies, Craik and Lockhart (1972) suggested that handling of physical and sensory stimulus such as lines, angles, pitch, and loudness is processed at the shallow level while recognizing patterns, finding the appropriate meaning of the word, and matching the new information with past learning are processed at the deeper level.

According to the Depth of Processing Hypothesis, semantic processing of lexical items results in higher retention than phonological or orthographical processing. Particularly, when a new word triggers associations, images or stories based on a person's past experience, the retention of the word can be strengthened (Craik & Lockhart, 1972).

The Depth of Processing Hypothesis, however, was criticized and eventually even "abolished" (Laufer & Hulstijn, 2001, p. 5) because of several critical problems. This hypothesis had a crucial limitation in that it did not provide any clear explanation of what constitutes, and how to measure the level of processing (Baddeley, 1978; Laufer & Hulstijn, 2001). Anderson (1979) stated that the Depth of Processing Hypothesis somewhat relies on "subjective intuitions" (p. 386). He pointed out that since there exist no objective rules for measuring the *depth* of a task, the hypothesis lacks operational definition. However, despite these shortcomings, the value of the Depth of Processing Hypothesis has been acknowledged for its contribution to predicting the effects of a task at least to some extent.

Craik and Tulving (1975) expanded the Depth of Processing Hypothesis, suggesting that the notion of *elaboration* should be added to the previous model. They contended that "retention differences should be attributed to degrees of stimulus elaboration rather than to differences in depth" (p. 279). Here,

elaboration refers to the process of establishing links between newly encountered information and previously stored information. In other words, it means creating context by connecting the target word with additional words related to it. Craik and Tulving (1975) stated that when individuals process items in an elaborative fashion, memory is enhanced since it provides an alternative way to retain and retrieve the information.

The effects of elaboration have been supported by a number of other psychological studies (Anderson, 1979; Lawson & Hogben, 1996; Schmitt & Schmitt, 1995). According to Anderson (1979), "it is not the depth of processing per se that is important, but one's prior practice at making elaborations" (p. 390). He emphasized the effects of elaboration on learning, suggesting that the better processing is "that which generates more elaborations of the input" (p. 390). Schmitt and Schmitt (1995) also accentuated the importance of elaboration in enhancing memory, demonstrating that rich and numerous associations with existing knowledge increase the chances that the new information is retained. This finding was corroborated by Lawson and Hogben's (1996) study, which revealed that the learners who elaborated the target word within a network of meaning or context could be more successful in learning.

2.2.2 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), based on the Depth of Processing Hypothesis (Craik & Lockhart, 1972). 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, 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 context 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, 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'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.

2.3 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 (Craik & Tulving, 1975) 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, demonstrating that the knowledge associated with the self was better remembered than the knowledge learned through other types of elaboration (e.g., Bower & Gilligan, 1979; Brown, Keenan, & Potts, 1986; Holland & Kensinger, 2010; Macrae et al., 2004; Maki & McCaul, 1985; Reeder et al., 1987). In Bower and Gilligan's (1979) experiment, the participants were assigned one of the three orienting tasks to memorize the given phrases (e.g., *a boring lecture, a faithful pet*). One task required self-reference encoding (judging if the given phrase describes the participants), another task demanded semantic encoding (judging if the phrase refers to a social interaction), and the last required structural encoding (judging if there are two or more 'e's in the phrase). This experiment revealed the superiority of the self-reference task to the other two tasks in the recall of the

target items. This indicates that relating the target items to episodes from one's history is more effective in memory than other types of elaboration tasks.

In an attempt to examine the reference type more closely, 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., Brown, Keenan, & Potts, 1986; Holland & Kensinger, 2010; Macrae et al., 2004; Maki & McCaul, 1985; Reeder et al., 1987).

Macrae et al.'s (2004) and Reeder et al's (1987) experiments were different from the previous studies in terms of the methodology; yet, their findings still corroborated the effects of autobiographical elaboration on memory. Macrae et al. (2004) assigned the autobiographical elaboration task to all the participants, and asked them to answer with "Yes" or "No" to the question "Does this adjective describe you?". As a result of analyzing the scores in the surprise word recall test,

the items related to the self were found to be better recalled than the counterparts that were not related to the self.

In Reeder et al.'s (1987) study, the experimental group was instructed to read a prose consisting of about 100 words, making a judgment whether the story was related to the participants themselves. In contrast, the comparison group was asked to think whether the prose was related to Princess Diana. After reading the prose, the participants were asked to retell the whole story. The result revealed that the task relating the prose to the self was more effective than the task associating the text with Princess Diana in remembering the story.

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 *self-memory system* (2000), 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. Thus, the current study compared two types of sentence-writing tasks: the autobiographical sentence-writing task (writing about the self) and the imaginary sentence-writing task (writing about an imaginary person). The latter task was used as the comparison task for the autobiographical sentence-writing task.

In addition to autobiographical elaboration, other aspects of the sentences such as length, lexical properties, and syntactic properties were also examined since there have been no attempts to examine the quality of the sentences. Considering the possibility that the differences in the quality of the sentences may affect vocabulary gains, the sentences written by the learners were closely examined in the present study.

CHAPTER 3
METHODOLOGY

In this chapter, the methodology of the current research will be discussed in terms of the participants, instruments, assessment, procedure, and data analysis.

3.1 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; Knight, 1994; Lee, 2006; Prince, 1996; Pulio, 2003; Sung, 2013, Swanborn & De Glopper, 2002; Van Daalen-Kapteijns et al., 2001; 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 3.1
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

3.2 Instruments

The instruments consist of target words and three tasks. They will be described in the following subsections respectively.

3.2.1 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. For example, in the case of the verb *deign*, its meaning is 'to do something reluctantly that one considers to be below one's dignity or standard', and there is no Korean word corresponding to this word. Like *deign*, those regarded as hard to translate into Korean were excluded from the target vocabulary.

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*).

3.2.2 Tasks

The sentence-writing task and the gap-filling task were conducted; as for the sentence-writing task, two types of tasks were provided for the students: the autobiographical sentence-writing task and the imaginary sentence-writing task.

3.2.2.1 Gap-filling Task

The participants in the gap-filling task group were asked to fill in the blanks in the given sentences (See Appendix 1). 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.

3.2.2.2 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 (See Appendix 2). 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.

3.2.2.3 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 (See Appendix 3). 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.

3.3 Assessment

The learners' active word learning and passive word learning on the target words were measured in both the immediate and the delayed tests. The methods for measuring the participants' active and passive word learning are described in this section.

3.3.1 Active Word Learning Test

In the active word learning test (see Appendix 4 for the immediate active test and Appendix 6 for the delayed active 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).

Based on the lexical production scoring protocol, the score was given according to the percentage of *correct* letters (letters placed in the correct positions) and the percentage of *present* letters (letters not placed in the correct positions). For example, as for the word *squabble*, the answers '…quabble' was awarded 0.75 points since more than 50 percentage (87.5%) were placed in their correct positions, and 'babble' was given 0.50 points because more than 50 percentage (62.5%) were present but in wrong positions. More detailed criteria of the lexical production scoring protocol are presented in Table 3.2. Scoring was conducted by the researcher of the study and an English teacher who had taught English for 4 years in middle and high schools. 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 the data analysis.

TABLE 3.2
Lexical Production Scoring Protocol-Written (Barcroft, 2002)

Points	Description
0.00	None of word is written; this includes - Nothing is written. - The letters present do not meet any "for 0.25" criteria.
0.25	1/4 of word is written; this includes - Any 1 letter is correct. - Correct number of syllables is less than 25%. - 25-49.9% of the letters are present.
0.50	1/2 of word is written; this includes - 25-49.9% of the letters are correct. - 50-74.9% of the letters are present.
0.75	3/4 of word written; this includes - 50-99.9% of the letters are correct. - 75-100% of the letters are present. - 100% of the letters are correct but other letters are added.
1.00	Entire word is written; this includes - 100% of the letters are correct.

3.3.2 Passive Word Learning Test

In the passive word learning test (See Appendix 5 for the immediate passive test and Appendix 7 for the delayed passive 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. For example, as for the adjective *insipid*, its Korean translation is *jiruhan,* but a student wrote its meaning as *jaemi umneun.* This answer triggered a little confusion about whether to give it a half point or the full point; through discussions, the students' score was fixed for 1.0 point since both raters agreed that *jiruhan* and *jaemi umneun* have almost the same meaning. 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.

3.4 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.

3.4.1 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 *irk*ed 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.

3.4.2 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.

3.5 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 with respect to the effects of the autobiographical sentence-writing task and the imaginary sentence-writing task on vocabulary learning, 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.

As for the last research question, that is, whether there are any differences between the sentences written by the learners who performed better in vocabulary retention and the sentences written by those who showed less vocabulary gains, the length, the lexical properties, and the syntactic properties of the sentences were closely examined. The sentences for examination were written by 20 students whose total scores of the delayed active and the delayed passive tests were the highest and 20 students whose total scores of the two delayed tests were the lowest. The 20 students with the intermediate scores were excluded to enlarge the gap between the two score groups. The delayed post-test scores were chosen as the criteria for selecting data since the purpose for this analysis is to investigate the factors related to long-term retention rather than initial learning.

The sentence analysis was conducted using the web-based *Coh-Metrix 3.0*., which is a computational tool to analyze the linguistic and discourse representations of the text (Coh-Metrix, 2015). This tool, which was developed at the University of Memphis, is mainly utilized for the investigation of cohesion or coherence of a written discourse; yet, it also includes some indices for examining the texts written at the sentence level.

Coh-Metrix has a shortcoming as a text analyzer in that it does not include some measures for lexical diversity and syntactic complexity; it uses only TTR (Type-Token Ratio), MTLD (Measure of Textual Lexical Diversity), and

vocd (Vocabulary Diversity) as the measures for analyzing the lexical diversity. When it comes to the syntactic complexity, it presents only the number of words before the verb and the number of modifiers per noun phrase; many other indices such as the number of AS-unit or the ratio of subordinate clauses are not included. Although there exist at least 19 measures for lexical diversity and 14 for syntactic complexity (Kim, 2014), many of them are excluded from *Coh-Metrix* indices. Despite this shortcoming, *Coh-Metrix* was selected as the text analyzer for the present study since it provides the most diverse information about written texts among the free online text analyzers. It displays various aspects of a written text with 108 indices. In addition, since the system is relatively stable, users rarely experience errors while using this software.

The whole number of the indices provided by *Coh-Metrix* is 108; yet, only 32 were selected for the current study since they were considered appropriate for the examination of the texts at the sentence level. The rest indices were excluded based on the researcher's judgment that they are less proper for analyzing individual sentences (e.g., *the number of paragraphs, the mean length of paragraphs*), and what they measure is overlapped with what other indices measure (e.g., *hypernymy for nouns and verbs*, which just combines other two indices, *hypernymy for nouns* and *hypernymy for verbs*). The selected indices are

presented in Table 3.3; according to their common properties, they are divided into three categories, that is, length, lexical properties, and syntactic properties.

TABLE 3.3
***Coh-Metrix* Indices Used for the Current Study**

Category	Index
Length	# of words
Lexical Properties	***Lexical Diversity***
	TTR for content words
	TTR for all words
	MTLD
	vocd
	Word Information
	# of syllables in a word
	# of letters in a word
	Noun incidence
	Verb incidence
	Adjective incidence
	Adverb incidence
	Pronoun incidence
	Word frequency of content words
	Word frequency of all words
	Polysemy for content words
	Hypernymy for nouns
	Hypernymy for verbs
	Connective Incidence
	Causal
	Logical
	Adversative and contrastive

	Temporal
	Additive
	Syntactic Complexity
	# of words before the main verb
	# of modifiers
	Syntactic Pattern Incidence
Syntactic Properties	NP
	VP
	AdvP
	PP
	Agentless passive voice
	Negation
	Gerund
	Infinitive

Note. Incidence means the number of occurrences per 1,000 words.

CHAPTER 4
RESULTS AND DISCUSSION

This chapter presents the results of the statistical analysis and discusses the findings. It is divided into three sections according to each of the three research questions as follows. Section 4.1 explores whether the sentence-writing task is more effective than the gap-filling task in the vocabulary learning of the learners at different proficiency levels. Section 4.2 discusses the effects of autobiographical elaboration in the sentence-writing task on the vocabulary learning of the learners at different proficiency levels. Lastly, Section 4.3 investigates the differences between the sentences written by the learners who achieved higher vocabulary retention and the sentences written by those who showed less vocabulary gains.

4.1 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 4.1 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 4.1
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. Table 4.2 shows the results.

TABLE 4.2
The Effects of the Task and the Proficiency on the Post-test Scores

Immediate Active Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	140.45	1	140.45	34.00	.000*	0.31
Proficiency	56.11	1	56.11	13.58	.000*	0.15
Task*Proficiency	2.11	1	2.11	0.51	.477	0.01
Error	313.96	76	4.13			
Immediate Passive Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	281.25	1	281.25	50.83	.000*	0.40
Proficiency	105.80	1	105.80	19.12	.000*	0.20
Task*Proficiency	0.45	1	0.45	0.08	.776	0.00
Error	420.55	76	5.53			
Delayed Active Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	45.75	1	45.75	26.08	.000*	0.26
Proficiency	17.11	1	17.11	9.76	.003*	0.11
Task*Proficiency	0.80	1	0.80	0.46	.502	0.01
Error	133.31	76	1.75			
Delayed Passive Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	130.05	1	130.05	31.08	.000*	0.29
Proficiency	68.45	1	68.45	16.36	.000*	0.18
Task*Proficiency	0.00	1	0.00	0.00	1.000	0.00
Error	318.05	76	4.18			

Note. *. The mean difference is significant at the 0.05 level.

As shown in Table 4.2, 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, 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, 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, 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.

4.2 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 4.3

displays the means and the standard deviations of the post-test scores of the learners.

TABLE 4.3
Descriptive Statistics of the Post-test Scores

Profi-ciency	Task	Immediate Test				Delayed Test				*N*
		Active		Passive		Active		Passive		
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
HP	Autobio-graphical	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	Autobio-graphical	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	Autobio-graphical	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, a set of two-way ANOVAs were conducted. The results are presented in Table 4.4.

As indicated in Table 4.4, no significant difference was found between the effects of the two types of sentence-writing tasks (all $p>.05$), which 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.

TABLE 4.4
The Effects of the Task and the Proficiency on the Post-test Scores

Immediate Active Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	4.56	1	4.56	0.75	.391	0.02
Proficiency	18.23	1	18.23	3.02	.091	0.08
Task*Proficiency	9.03	1	9.03	1.49	.230	0.04
Error	217.51	36	6.04			
Immediate Passive Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	0.00	1	0.00	0.00	1.000	0.00
Proficiency	46.23	1	46.23	5.82	.021*	0.14
Task*Proficiency	15.63	1	15.63	1.97	.169	0.05
Error	286.05	36	7.95			
Delayed Active Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	0.76	1	0.76	0.25	.620	0.01
Proficiency	5.26	1	5.26	1.74	.196	0.05
Task*Proficiency	0.51	1	0.51	0.17	.685	0.00
Error	108.80	36	3.02			
Delayed Passive Test						
	SS	*df*	*MS*	*F*	*p*	Partial η^2
Task	8.10	1	8.10	1.18	.284	0.03
Proficiency	34.23	1	34.23	5.00	.032*	0.12
Task*Proficiency	5.63	1	5.63	0.82	.371	0.02
Error	246.45	36	6.85			

Note. *. The mean difference is significant at the 0.05 level.

As for the interaction between the task and the proficiency, 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 result 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; Brown et al., 1986), 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.

4.3 Sentence Analysis

In order to detect the differences between the sentences written by the learners who attained higher vocabulary retention and the sentences written by those who showed less vocabulary gains, a series of *T*-tests were conducted on each of the 32 *Coh-Metrix* indices. The independent variable was the post-test score group: the high score group and the low score group. As described in the chapter 3, the high score group comprised 20 learners whose total scores of the delayed active and passive tests were the highest in the sentence-writing group, and the low score group consisted of 20 learners whose total scores of the two delayed tests were the lowest. The descriptive statistics and the *T*-test results are

presented in Table 4.5.

TABLE 4.5
***T*-test on the *Coh-Metrix* Results**

Variable	High Score Group		Low Score Group		$T(1, 38)$
	M	*SD*	*M*	*SD*	
LENGTH					
# of words	102.00	23.22	83.80	24.96	.022*
LEXICAL PROPERTIES					
Lexical Diversity					
TTR for content words	0.74	0.09	0.71	0.14	.545
TTR for all words	0.61	0.05	0.59	0.08	.415
MTLD	47.83	18.08	35.82	18.68	.046*
vocd	25.85	27.34	13.28	22.90	.124
Word Information					
# of syllables in a word	1.35	0.06	1.35	0.08	1.000
# of letters in a word	4.02	0.24	3.98	0.32	.656
Noun incidence	297.38	66.56	348.92	132.44	.131
Verb incidence	167.91	32.73	161.08	38.05	.546
Adjective incidence	66.28	20.85	71.42	18.81	.418
Adverb incidence	57.31	17.47	43.38	23.03	.038*
Pronoun incidence	161.96	73.03	146.57	97.13	.575
WF of content words	2.49	0.15	2.51	0.26	.726
WF of all words	2.99	0.12	2.94	0.27	.426
Polysemy for content words	4.08	0.75	3.91	0.52	.409
Hypernymy for nouns	7.31	1.21	7.28	1.00	.919
Hypernymy for verbs	1.37	0.23	1.33	0.20	.566
Connective Incidence					

Causal	63.31	27.72	45.78	24.69	.041*
Logical	56.08	25.73	48.37	24.15	.335
Adversative and contrastive	2.35	4.23	0.43	1.94	.077
Temporal	9.36	10.60	11.67	18.17	.625
Additive	10.80	11.02	6.40	11.76	.230
SYNTACTIC PROPERTIES					
Syntactic Complexity					
# of words before main verb	1.43	0.33	1.45	0.54	.892
# of modifiers per noun phrase	0.41	0.10	0.44	0.19	.589
Syntactic Pattern Incidence					
NP	416.13	26.01	425.28	50.37	.477
VP	244.26	19.79	242.47	43.36	.868
AdvP	18.63	10.57	12.82	12.13	.115
PP	48.88	21.20	51.10	26.35	.771
Agentless passive voice	2.98	5.67	3.47	8.04	.826
Negation	19.90	14.46	8.93	10.48	.009*
Gerund	13.39	12.62	8.08	10.27	.153
Infinitive	6.13	10.61	4.06	13.75	.596

Note. *. The mean difference is significant at the 0.05 level. # = number; WF = word frequency; NP = noun phrase; VP = verb phrase; AdvP = adverbial phrase; PP = preposition phrase

The statistics results will be discussed with regard to three properties: length, lexical properties, and syntactic properties. Each of them will be discussed in the following three subsections.

4.3.1 Length

As for the length of the sentences, the number of words was calculated. *Coh-Metrix* provides the information of the whole text written with the 10 target words. Considering this, the high score group and the low score group wrote 10.2 and 8.38 words for each target word respectively. This indicates that on average, the high score group produced 1.82 words longer context for each target word than the low score group; the difference between the two groups was found statistically significant ($p<.05$). With regard to this finding, there have been some empirical studies which demonstrated that the length of context affects the memory of the target words. For example, Rohwer (1968) suggested that the greater the elaboration, the more learning efficiency increases. This argument was confirmed by Turnure's (1971) study, which demonstrated that the learners who read a paragraph containing the target word performed better in the recalling test than those who were provided with only one sentence including the target word. These findings indicate that the length of context may affect the learning efficiency of new words.

4.3.2 Lexical Properties

The lexical properties of the sentences were analyzed with regard to lexical diversity, word information, and the connective incidence. First, as for the lexical diversity, TTR for content words, TTR for all words, MTLD, and *vocd*

were used as the measures.

As a result of comparing the lexical diversity of the sentences written by the two score groups through *T*-tests, a significant difference was found in MTLD ($p<.05$). Although TTR is the best-known measure for lexical diversity, it has a critical shortcoming in that it is easily affected by the length of the text; in this sense, MTLD provides more reliable results because it is not influenced by the length of the text (McCarthy & Jarvis, 2010; McNamara & Graesser, 2012). *Vocd* is not affected by the length of the text as well; yet, because of its sampling method, it tends to fluctuate when lexical diversity is very high (McCarthy & Jarvis, 2010). Considering the high TTR value in the current study (higher than 70% for content words), MTLD may provide the most reliable result among the measures used for the current study.

Based on the MTLD values, the learners who showed better retention of the target words used significantly more diverse words in formulating sentences. This result can be explained by Craik and Tulving's (1975) claim that when an alternative way to retain and retrieve the word is provided, the retention of the target word is enhanced. This implies that when learners use diverse words in making sentences, they have more alternative cues to remember the target word. The following examples present the difference between the sentences written by the two score groups.

(1) Tom *mar*ed his dinner since he had omited ingredients. (High score group)

(2) I exam past week and I very *mar* exam past week. (Low score group)

Although the sentence (2) includes more words than the sentence (1), it does not include diverse words. The writer of the sentence (2) tends to repeat the same words or phrases, such as *I, exam,* and *past week.* In contrast, more various words are found in the sentence (1), and there is no repeated word or phrase in this sentence. The lexical diversity of the sentence (1) seems to add more abundant context to the target word *mar.* Unlike the sentence (2), the sentence (1) involves more specific information such as the reason why Tom failed to prepare a nice meal by adding the subordinate clause, *since he had omitted ingredients*. Thus, the writer of the sentence (1) may have had more clues in retrieving the target word than the student who wrote the sentence (2).

In the next category, *word information*, a significant difference was found in the frequency of the adverb. The high score group was found to use more adverbs than the low score group, as indicated in the following examples.

(3) Tom *mar*ed the relationship of his brother and him. He beated his brother **so much**. (High score group)

(4) Tom is *mar* math because he don't like math. (Low score group)

As shown in the sentences above, the student of the high score group describes the situation more vividly by using the adverb *so much* while the student of the low score group does not clarify how much Tom does not like math. As in the sentence (3), the intensifiers such as *very, so, too, really*, or *much* were more frequently found in the high score group's sentences than in the low score group's sentences. A few more types of adverbs that were frequently used by the high score group are presented below.

(5) The dog *irk*ed for Tom because the dog barked **loudly**. (High score group)

(6) **Yesterday**, one student splashed food sauce. She was *surly*. (High score group)

The sentence (5) shows the use of the manner adverb *loudly.* By using this manner adverb, the learner describes the dog's behavior more vividly. In terms of the sentence (6), the learner uses the temporal adverb *yesterday*, specifying the temporal setting of the event. These types of adverbs were not frequently found in the sentences written by the low score group. From this finding, it can be concluded that the high score group tends to make the context

more specific and detailed by using intensifiers, manner adverbs, or temporal adverbs.

In the next category, the *connective incidence*, no significant difference was found in relation to most connective types except for the causal connective. This result may be attributed to the fact that the sentence-writing task in the current study required writing only one or two sentences. In this condition, the students may have had fewer opportunities to use connectives. Yet, interestingly, a significant difference was found in the causal connective incidence. When looking at the sentences written by each group, causal connectives were found more frequently in the high score group's sentences. See the following examples:

(7) Tom did big *bungle* for this English cession.

Tom hate Tom's *chum* Jerry **because** Jerry is surly.

Tom is *insipid* **because** he watching boring movie.

The dog *irk*ed for Tom **because** the dog barked loudly.

Tom *mar*ed his vacation plan **because** his father was died.

Tom has *prowess*. For example, to fly, to swim, and so on.

Tom *squabble* his friend **because** his friend forget to bring Tom's money.

Tom is *surly* **because** he fighted man.

Tom *upbraid* his brother **because** his brother didn't homework.

Tom is *voluble*. **So** he easy to friend. (High score group)

(8) I did *bungle* the test. That is important in yesterday.

My close *chum* is kind and handsome.

I think *insipid* that I study English.

He *irk* me when he touch my body.

I *mar* test when each years.

My *prowess* is that sing a song well.

I hate *squabble*. That is **because** I fighted close friend.

I see *surly* those who he hit her.

I *upbraid* he that hit her.

My friend is *voluble*. He is nice. (Low score group)

As shown in the examples above, the learner of the high score group tends to use more causal connectives such as *because* and *so* while they are seldom found in the sentences written by the learner of the low score group. When looking into the sentences written by the learners of the low score group, they seem to depend more on the relative pronouns or temporal connectives in extending context, or write just one single clause without any further extension of the context.

The reason that the writer of the example (7) showed better performance in vocabulary learning can be explained by a number of studies conducted in the field of psychology (e.g., Black & Bern, 1981; Black & Bower, 1980; Bradshaw & Anderson, 1982; Graesser & Haberlandt, 1993; Myers et al., 1984; Trabasso & van den Broek, 1985). For example, Black and Bower (1980), and Trabasso and van den Broek (1985) revealed that the proximity to the causal chain was a powerful predictor of recall for story events. Likewise, Bradshaw and Anderson (1982) found that recall was better for a target sentence when it was studied with additional sentences causally related to it than when studied with unrelated sentences. Another interesting result was gained in Black and Bern's (1981) study, which demonstrated that the recall for causally related sentence pairs was better than that for temporally related pairs. This finding was corroborated by several subsequent studies (e.g., Myers et al., 1984; Millis, Graesser, & Haberlandt, 1993).

There have been more studies particularly focused on the effects of the conjunction on recalling the sentences (Caron, Micko, & Thuring, 1988; Townsend, 1983). For example, in Caron, Micko, and Thuring's (1988) study, it was found that the sentences were better recalled when the sentence pair was linked by the connective *because* than when the sentence pair was unconnected or connected by *and* or *but*. From this finding, the researchers concluded that the conjunction *because* induces the stronger association between the sentences,

compared to other conjunctions such as *and* or *but*. According to them, the latter conjunctions are less effective in constructing a coherent mental representation integrating the sentences, which leads to the low rate of retrieval of the target sentence.

Based on the previous studies, the context including the causal relatedness or causal connectives is more likely to be recalled than the context without causality or that including other types of connectives such as temporal (e.g., *when*), additional (e.g., *and*), or adversative (e.g., *but*) connectives. This implies that the causality of the context is conducive to enhancing memory.

4.3.3 Syntactic Properties

The syntactic properties of the sentences were analyzed based on the syntactic complexity and the syntactic pattern of the sentences. First, as for the syntactic complexity, *Coh-Metrix* results presented two indices: the number of words before the main verb and that of modifiers per noun phrase. With regard to both indices, the two score groups did not show any significant difference ($p>.05$), which indicates that writing syntactically complex sentences does not affect vocabulary learning.

Next, in terms of the syntactic pattern incidence, a significant difference was found in the incidence of negation markers ($p<.05$). With regard to this

finding, Kensinger and Corkin's (2003) study provides some explanations. In their research, the participants could remember negative information better than neutral information. This implies the possibility that the negative content can be remembered better than the neural content. Another possible explanation is that since most of the target words (*bungle, insipid, irk, mar, squabble, surly*, *upbraid*) except for only three (*chum, prowess*, *voluble*) were negative words, the high score group's sentences contained a number of negative words. The sentences (9) and (10) below exhibit the characteristics of the sentences written by the high score group.

(9) Tom have *bungle* in the test. He didn't mark to OMR. (High score group)

(10) I had a *squabble* with Hyebin. Because she didn't like my attitude. (High score group)

As shown in (9) and (10), the students of the high score group used the negation marker *didn't* to extend the context about the negative word *bungle* and *squabble.* In contrast, most learners of the low score group used fewer negation markers since they added less context to the target word. The sentences written by the low score group are presented below.

(11) I *squabble* my father about my dream. (Low score group)

(12) Tom make *bungle* when he play basketball. (Low score group)

Since the low score group seldom creates long context for the target word, there seem to be fewer opportunities for them to use the negation markers.

To sum up, the differences between the high score group and the low score group were found in the length of context, lexical diversity, and the frequencies of adverbs, causal connectives, and negation markers. This implies that the learners who show better vocabulary retention tend to write longer context, using more various words. They seem to make their context more specific and vivid by adding more adverbs such as intensifiers, manner adverbs, and temporal adverbs. In addition, they tend to strongly relate two clauses by inserting causal connectives such as *because* and *so*. These causal connectives have been proved to have a positive effect on the recall of information. In terms of the negation markers, the high score group was found to use them more frequently than the low score group. With regard to this result, two explanations can be provided. First, the negative content may be remembered better than the neural content. Second, since most of the target words were negative words, the high score group learners, who tended to produce long sentences, may have used more negative expressions in extending context.

CHAPTER 5
CONCLUSION

This chapter consists of three sections. In Section 5.1, the major findings of the current study are summarized, and in Section 5.2, pedagogical implications are provided. Lastly, Section 5.3 presents the limitations of the present study and suggestions for further research.

5.1 Major Findings

The current study explored the effects of the sentence-writing task on English vocabulary learning of Korean high school learners at different proficiency levels. The first research investigated whether the sentence-writing task is more effective than the gap-filling task on vocabulary learning. In the second research, the autobiographical sentence-writing task and the imaginary sentence-writing task were compared in terms of their effects on vocabulary learning. The third research explored the differences between the sentences written by the learners who performed better in vocabulary retention and the sentences written by those who showed less vocabulary gains with regard to length, lexical properties, and syntactic properties.

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 proficiency 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, 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; Brown, Keenan, & Potts, 1986; 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

may yield almost the same results as the autobiographical sentence-writing task on vocabulary learning.

When it comes to the third research question, the current study explored whether there is any difference between the sentences written by the learners who performed better in vocabulary retention and the sentences written by those who showed less vocabulary gains. A series of *T*-tests were conducted between the high score group and the low score group, using the 32 *Coh-Metrix* indices as the dependent variables. From the *T*-test results, it was concluded that the sentences written by the two score groups were significantly different in length, lexical diversity, and the frequencies of adverbs, causal connectives, and negation markers. In other words, the learners of the high score group created longer context with more various words, using more adverbs, causal connectives, and negation markers. With regard to these results, some previous studies in the field of psychology can provide some explanations. Rohwer's (1968) and Turnure's (1971) studies revealed that the length of context affects the learning efficiency of target items. When it comes to lexical diversity, Craik and Tulving (1975) claimed that when an alternative way to retain and retrieve the word is provided, the retention of the target word is enhanced. This implies that when learners use diverse words in creating sentences, they have more alternative cues to remember the target word.

Next, as for the adverb incidence, the high score group was found to use more intensifiers, manner adverbs, and temporal adverbs, which made the described situation more vivid and detailed. In terms of the causal connective, there have been a number of studies that confirmed the positive effects of causal cohesion on memory (e.g., Black & Bern, 1981; Black & Bower, 1980; Bradshaw & Anderson, 1982); thus, it may be helpful in vocabulary learning to include causal cohesion in sentences.

Lastly, when it comes to negative markers, they were found to be more frequently used by the high score group than by the low score group. This result can be interpreted based on Kensinger and Corkin's (2003) finding, which revealed that negative information can be remembered better than neutral information. However, there is another possibility that the high score group used many negative markers in extending context since most of the target words were negative vocabulary.

5.2 Pedagogical Implications

Based on the major findings presented in Section 5.1, the present study provides the following pedagogical implications on English vocabulary learning. First, assigning the sentence-writing task to learners may be more effective in their vocabulary learning than providing them with the gap-filling task. The

sentence-writing task was found to be significantly more effective than the gap-filling task in both the short-term and the long-term learning of new vocabulary, regardless of learners' proficiency levels. The effects of the sentence-writing task can be observed more clearly when the content guideline, which is conducive to extending the length of the sentences, is provided.

Second, the autobiographical sentence-writing task and the imaginary sentence-writing task yield almost the same outcomes in vocabulary learning. Even if the learners are assigned the imaginary sentence-writing task, they tend to depend on their own past experiences in creating sentences; thus, the boundary between the autobiographical sentence-writing and the imaginary sentence-writing is somewhat blurred. Therefore, it may be more desirable for the teacher to encourage the students to create sentences in whichever way that makes it easier for them or motivates them to do the tasks.

Third, when practitioners provide the sentence-writing task for learners, it may be important to improve the quality of their sentences particularly in terms of length, lexical diversity, and the use of adverbs and causal connectives. By presenting appropriate content guidelines, the quality of the learners' sentences can be improved. For example, rather than just letting learners write sentences freely, presenting more specific content guidelines such as "Write about your own experience where you had a *squabble* with someone else. Please specify with

whom you fought and why." could be helpful in extending the context for the target word, including causal cohesion. Improving the sentence quality in this way might enhance the learners' memory of the target words.

5.3 Limitations and Suggestions for Further Research

This section provides the limitations of the current study and some suggestions for further 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 the word knowledge is. For example, Nation (2001) suggested that if one is to be regarded to have full knowledge of a certain word, at least the knowledge of *form* (pronunciation and spelling), *meaning*, and *use* (grammatical functions, collocations, and constraints on use such as register) is necessary. Given that acquiring vocabulary involves such diverse aspects of a word, the vocabulary tests used in the current study might not be sufficient to measure the depth of learners' vocabulary knowledge. In order to measure other aspects such as the knowledge of grammatical functions, collocations, or register, additional tests should be focused on the capacity to use the target word in appropriate context. Moreover, to measure the collocational knowledge, it would be necessary to provide learners

with sufficient information about collocations. It would be possible to select collocations as the target vocabulary instead of individual words as well. In addition, when it comes to the knowledge about the constraints on use, only one exposure to the target word may not be sufficient to gain such knowledge; thus, multiple exposures in diverse context would be necessary.

Another limitation of the present study is related to the sentence analysis. Although the results of the sentence analysis revealed that certain properties were the main differences between the sentences written by the high score group and those written by the low score group, it does not necessarily mean that such properties (length, lexical diversity, the frequencies of adverbs, causal connectives, and negation markers) significantly affect vocabulary learning. In order to confirm whether these variables are the significant factors in vocabulary learning, further experimental research should be designed with each property as the independent variable and learners' vocabulary gains as the dependent variable.

There is another limitation of the current research; that is, the study was conducted only with the students from one specific high school; accordingly, there is a limitation to generalize this finding to other high school students. To improve the reliability of the research, it is recommended that future research be conducted with the students from several different high schools. In addition, given that high school students have spent much more time on studying English than middle

school students, the research conducted with middle school students may yield different results. Therefore, it is recommended to conduct further research with middle school students as well as high school students.

REFERENCES

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. *Levels of Processing in Human Memory*, *104*, 385-403.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, *85*(3), 139-152.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*(2), 323-363.

Black, J. B., & Bower, G. H. (1980). Story understanding as problem-solving. *Poetics*, *9*(1), 223-250.

Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 267-275.

Bobrow, S. A., & Bower, G. H. (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology*, *80*(3), 455-461.

Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality*, *13*(4), 420-432.

Bradshaw, G. L., & Anderson, J. R. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 165-174.

Brown, H. D. (2007). *Principles of language learning and teaching*. Cambridge: Cambridge University Press.

Brown, P., Keenan, J. M., & Potts, G. R. (1986). The self-reference effect with imagery encoding. *Journal of Personality and Social Psychology*, *51*(5), 897-906.

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.

Coh-Metrix version 3.0 indices. (2015). (**http://cohmetrix.com**)

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words

in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268.

Day, R. R., & Swan, J. (1998). Incidental learning of foreign language spelling through targeted reading. *TESL Reporter*, *31*(1), 1-9.

de Groot, A. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning, 56*(3), 463-506.

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.

Elley, W. B., & Mangubhai, F. (1981). The long-term effects of a book flood on children's language growth. *Directions*, *7*, 15-24.

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

Grabe, W., & Stoller, F. L. (1997). Reading and vocabulary development in a second language. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 98-122). Cambridge: Cambridge University Press.

Hafiz, F. M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT Journal*, *43*(1), 4-13.

Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of Life Reviews*, *7*(1), 88-131.

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., & 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.

Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System*, *18*(2), 199-207.

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words?. *Memory & Cognition*, *31*(8), 1169-1180.

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal, 78*(3), 285-299.

Kim, H. S., & Na, Y. H. (2010). Vocabulary learning and task-induced involvement. *Applied Linguistics, 26*(4), 183-211.

Kim, J. Y. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching, 69*(4), 27-51.

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

Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla Yearbook*, *5*(1), 223-250.

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning*, *46*(1), 101-135.

Lee, Y. H. (2006). *The impact of task-induced involvement load on EFL*

vocabulary learning. Unpublished master's thesis. Seoul National Universtiy.

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.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381-392.

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 188-205). Hershey, PA: IGI Global.

Millis, K. K., Graesser, A. C., & Haberlandt, K. (1993). The impact of connectives on the memory for expository texts. *Applied Cognitive Psychology*, *7*(4), 317-339.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56-60.

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*(5), 519-533.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*(4), 453-465.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253.

Nation, I. S. P., & Ming-Tzu, K. W. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*(2), 355-380.

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.

Norman, D. A. (1969). *Memory and attention*. New York: Wiley.

Park, J. Y., & Oh, J. I. (2015). Effects of task-induced vocabulary learning invovlement on vocabulary learning by Korean university students. *English Language Teaching, 27*(3), 175-191.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart?. *Reading in a Foreign Language*, *22*(1), 31-55.

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.

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *Modern Language Journal*, *80*(4), 478-493.

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, *53*(2), 233-284.

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., et al. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*(9), 677.

Rohwer Jr, W. (1968). Mental mnemonics in early learning. *The Teachers College Record*, *70*(3), 213-226.

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: Theoretical underpinnings and practical suggestions. *ELT Journal*, *49*(2), 133-143.

Schmitt, N. (2008). Instructed second langauge 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.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition*

(pp. 89-105). Hillsdale, NJ: Erlbaum.

Sung, H. M. (2013). Task-induced involvement load in Korean EFL incidental vcabulary learning. *Language Research*, *29*(2), 269-296.

Swanborn, M. S. L., & De Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning*, *52*(1), 95-117.

Turnure, J. E. (1971). Types of verbal elaboration in the paired-associate performance of educable mentally retarded children. *American Journal of Mental Deficiency*, *76*(3), 306-312.

Thurston, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, *17*(3), 267-280.

Townsend, D. J. (1983). Thematic processing in sentences and texts. *Cognition*, *13*(2), 223-261.

Trabasso, T., & Van Den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612-630.

Van Daalen-Kapteijns, M., Elshout-Mohr, M., & De Glopper, K. (2001). Deriving the meaning of unknown words from multiple contexts. *Language Learning*, *51*(1), 145-181.

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.

Appendix 1

Gap-filling Task

■ **Fill in the blanks with appropriate words in the wordlist.**

1	He had never done that before. He made a lot of _____s due to his lack of experience. *due to: ~ 때문에 / *lack: 부족
2	Rachel was very tired, but the rehearsal seemed endless. The repeated practice _____ed her increasingly. *rehearsal: (공연 등의) 리허설 / *increasingly: 점점
3	Michael's son is a football player. Michael always brags about his son's sporting _____. *brag: 자랑하다 / *sporting: 스포츠의
4	David spent the whole weekend watching TV or sleeping. So his mother _____ed him about his laziness. *spend 시간 ~ing: ~하면서 시간을 보내다
5	Chang-ho is very _____ and sociable. He is good at making friends. *sociable: 사교적인
6	Min-hee's best _____ is Hye-ran. They first met in kindergarten, and they have known each other for ten years. *kindergarten: 유치원
7	I don't want to go to the store. The clerk in the store always behaves in a/an _____ manner to customers. *clerk: 가게 점원 / *manner: 태도
8	Emma went to see a musical yesterday. She was looking forward to seeing an exciting performance, but it was _____ from the beginning. *look forward to: ~을 기대하다

9	My brother messed up my room, so I had a/an _____ with him last night. I'm still angry at him. *mess up: 어지럽히다
10	More than 200 visitors came to enjoy the festival. But some unexpected problems _____ed the festival. *unexpected: 예상치 못한

Wordlist

bungle: 실수 (명사)	**chum**: 친구 (명사)
dissent: 반대하다 (동사)	**insipid**: 지루한 (형용사)
irk: 짜증나게 하다 (동사)	**mar**: 망치다 (동사)
prowess: 재주 (명사)	**remarkable**: 뛰어난 (형용사)
squabble: 다툼 (명사)	**surly**: 무례한 (형용사)
upbraid: 나무라다 (동사)	**voluble**: 입담이 좋은 (형용사)

■ **Finishing Time: _____**

Appendix 2

Autobiographical Sentence-writing Task

■ **Write one or two sentences with each given word. According to the content guidelines, please write about yourself. Please make sure to write your own experiences or thoughts. You do not have to care too much about grammatical or orthographical accuracy here.**

□ **bungle**: 실수 (noun)

※ Write about your own experience where you made a *bungle*. Please describe when and what *bungle* you made.

☞ _____

□ **chum**: 친구 (noun)

※ Write about your best *chum*. Please write his or her name and the reason why you like the person.

☞ _____

□ **insipid**: 지루한 (adjective)

※ Write about something that you think is *insipid*. Please explain the reason why you think it is *insipid*.

☞ _____

□ **irk**: 짜증나게 하다 (verb)

※ Write about your own experience where someone or something *irk*ed you. Please specify who/what *irk*ed you and how.

☞ _____

□ **mar**: 망치다 (verb)

※ Write about your own experience where you *mar*red something. Please specify what you *mar*red and why.

☞ _____

□ **prowess**: 재주 (noun)

※ Please write about something you have *prowess* in.

☞ _____

□ **squabble**: 다툼 (noun)

※ Write about your own experience where you had a *squabble* with someone else. Please specify with whom you fought and why.

☞ _____

□ **surly**: 무례한 (adjective)

※ Write about your own experience where you saw a *surly* person. Please describe how the person behaved in the *surly* way and towards whom.

☞ _____

□ **upbraid**: 나무라다, 비난하다 (verb)

※ Write about your own experience where you *upbraid*ed someone else or other people *upbraid*ed you. Please specify who *upbraid*ed whom and why.

☞ _____

□ **voluble**: 입담이 좋은 (adjective)

※ Write about your acquaintance who is *voluble*. Indicate who this person is and how you came to know the person.

☞ _____

■ **Finishing Time: _____**

Appendix 3

Imaginary Sentence-writing Task

■ **Write one or two sentences with each given word. According to the content guidelines, please write about an imaginary character Tom. You do not have to care too much about grammatical or orthographical accuracy here.**

□ **bungle**: 실수 (noun)

※ Write one or two sentences that illustrate a situation where Tom made a *bungle*. Please imagine and describe when and what *bungle* he made.

☞ _____

□ **chum**: 친구 (noun)

※ Write one or two sentences about Tom's *chum*. Please imagine and write his or her name and the reason why Tom likes the person.

☞ _____

□ **insipid**: 지루한 (adjective)

※ Write one or two sentences where Tom thinks something is *insipid*. Please imagine and write about what he thinks is *insipid* and why.

☞ _____

□ **irk**: 짜증나게 하다 (verb)

※ Write one or two sentences that illustrate a situation where someone or something *irk*ed Tom. Please imagine and describe who/what *irk*ed him and how.

☞ _____

□ **mar**: 망치다 (verb)

※ Write one or two sentences that illustrate a situation where Tom *mar*red something. Please imagine and write about what he *mar*red and why.

☞ _____

□ **prowess**: 재주 (noun)

※ Write one or two sentences that illustrate where Tom shows *prowess*. Please imagine and describe what he has *prowess* in.

☞ _____

□ **squabble**: 다툼 (noun)

※ Write one or two sentences that illustrate a situation where Tom had a *squabble*. Please imagine and describe with whom he had a *squabble* and why.

☞ _____

□ **surly**: 무례한 (adjective)

※ 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.

☞ _____

□ **upbraid**: 나무라다, 비난하다 (verb)

※ Write one or two sentences that illustrate a situation where Tom *upbraid*ed someone. Please imagine and write about whom he *upbraid*ed and why.

☞ _____

□ **voluble**: 입담이 좋은 (adjective)

※ Write one or two sentences where Tom is a kind of *voluble* person. Please imagine and describe how this personality is beneficial to him.

☞ _____

■ Finishing Time:_____

Appendix 4

Immediate Active Test

■ Write the English words corresponding to the given Korean words. If you cannot remember the whole spelling of the word, please write just some parts of the word that you can recall.

□ 지루한:

□ 다툼:

□ 나무라다:

□ 무례한:

□ 망치다:

□ 친구:

□ 실수:

□ 입담이 좋은:

□ 짜증나게 하다:

□ 재주:

Appendix 5
Immediate Passive Test

■ Write the Korean words corresponding to the given English words. Please write in the accurate form of the part-of-speech. [pretty: 예쁜(O), 예쁘다(△)]

□ surly:

□ bungle:

□ insipid:

□ chum:

□ upbraid:

□ mar:

□ voluble:

□ prowess:

□ irk:

□ squabble:

※ If there are some words that you already knew before this class, please check them in the following boxes.

□ bungle	□ chum	□ insipid	□ irk	□ mar
□ prowess	□ squabble	□ surly	□ upbraid	□ voluble

Appendix 6

Delayed Active Test

■ Write the English words corresponding to the given Korean words. If you cannot remember the whole spelling of the word, please write just some parts of the word that you can recall.

□ 망치다:

□ 재주:

□ 무례한:

□ 짜증나게 하다:

□ 지루한:

□ 입담이 좋은:

□ 다툼:

□ 나무라다:

□ 실수:

□ 친구:

Appendix 7

Delayed Passive Test

■ Write the Korean words corresponding to the given English words. Please write in the accurate form of the part-of-speech. [pretty: 예쁜(O), 예쁘다(△)]

□ voluble:

□ prowess:

□ squabble:

□ insipid:

□ irk:

□ surly:

□ mar:

□ upbraid:

□ bungle:

□ chum:

※ If there are some words that you reviewed after the previous class, please check them in the following boxes.

□ bungle	□ chum	□ insipid	□ irk	□ mar
□ prowess	□ squabble	□ surly	□ upbraid	□ voluble

국문초록

본 연구는 문장 쓰기 과제가 한국 고등학생들의 영어 어휘 학습에 미치는 영향을 알아보고자 수행되었다. 문장 쓰기 과제의 효과는, 관여도가 높은 과업이 학습자의 어휘 학습을 향상시킨다는 관여도 가설(Laufer & Hulstijn, 2001)을 바탕으로 설명될 수 있다. 많은 연구들이 관여도 가설을 검증하기 위해 수행되었지만, 연구들 간의 결과는 일치하지 않았다. 연구 결과의 불일치는 주로 문장 쓰기 과제와 빈칸 채우기 과제의 비교에서 나타났다. EFL 환경에서 이루어진 몇몇 연구들은 관여도가 높은 문장 쓰기 과제와 관여도가 낮은 빈칸 채우기 과제의 효과 간에 유의미한 차이가 없다는 결과를 보여왔다. 선행 연구에서 학습자 수준이 흔히 간과되었던 점을 고려하여, 본 연구에서는 수준이 서로 다른 학습자를 대상으로 문장 쓰기 과제와 빈칸 채우기 과제의 효과를 비교하였다. 또한 문장 쓰기 과제의 자서전적 정교화(어떤 어휘의 뜻을 자기 자신의 경험과 연결하는 것)가 어휘 학습에 미치는 효과를 다룬 연구가 없었기 때문에, 본 연구에서는 자서전적 문장 쓰기 과제의 효과와 상상적 문장 쓰기 과제의 효과를 비교하였다. 이 외에도, 높은 어휘 암기율을 보인 학습자와 낮은 어휘 암기율을 보인 학습자가 쓴 문장에 차이가 있는지를 살핀 연구가 없었던 점을 고려하여, 본 연구에서는 사후 시험 점수가 다른 학습자가 쓴 문장의 질을 서로 비교하였다.

본 연구에서는 경기도에 위치한 한 고등학교의 40명의 상위 학습자와 40명의 하위 학습자들이 참가자로 선정되어 문장 쓰기 과제와 빈칸 채우기 과제 중 한 가지 과제에 무작위로 배정되었다. 문장 쓰기 과제는 자서전적 문장 쓰기 과제와 상상적 문장 쓰기 과제로 나뉘어 수행되었고, 학생들이 문장의 내용을 생각해낼 수 있도록 짧은 가이드라인이 한국어로 제공되었다. 예를 들어, *surly*(한국어 뜻: 무례한)라는

단어의 경우, “자신이 무례한 사람을 보았던 경험에 대해 영어로 써봅시다. 그 사람이 누구에게, 어떻게 무례하게 행동했는지 적어보세요.”와 같은 가이드라인이 한국어로 제시되었다.

문장 쓰기 집단과 빈칸 채우기 집단 간에 이원분산분석을 실시한 결과, 관여도 가설과 일치하는 결과가 나타났다. 학습자의 수준에 관계 없이, 문장 쓰기 과제가 빈칸 채우기 과제보다 어휘 학습에 효과적인 것으로 나타났다. 이러한 결과가 나타난 데에는 문장 쓰기 집단에 주어진 내용 가이드라인이 영향을 미쳤을 가능성이 있다. 즉, 학습자들이 충분한 길이의 문장을 쓰는 데 있어서 구체적인 내용 가이드라인이 도움이 되었을 수 있다. 한편, 자서전적 문장 쓰기 과제와 상상적 문장 쓰기 과제의 효과 간에는 유의미한 차이가 나타나지 않았다. 이러한 결과는 학습자들이 자신의 과거 경험을 바탕으로 문장을 썼는지의 여부가 어휘 학습에 영향을 미치지 않는다는 것을 의미한다. 마지막으로, 학생들이 쓴 문장의 질적인 측면을 *Coh-Metrix 3.0*이라는 도구를 통해 분석한 결과, 사후 어휘 시험 점수가 높은 집단이 쓴 문장과 사후 어휘 시험 점수가 낮은 집단이 쓴 문장 간에는 맥락의 길이와 어휘적 다양성, 부사의 빈도, 인과적 접속사의 빈도, 그리고 부정어구의 빈도에 있어서 유의미한 차이가 나타났다. 이러한 결과는 학습자가 쓴 문장의 다양한 특성들이 어휘 학습에 영향을 미칠 수 있다는 점을 시사한다. 이러한 결과에 근거한 교육적 시사점들이 결론부에 제시되었다.

주요어: 어휘 학습, 문장 쓰기 과제, 관여도 가설, 자서전적 정교화

학 번: 2012-21372