



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학석사학위논문

The Effects of the Involvement Load of Tasks
on English Vocabulary Learning of
Korean High School Students

과업 관여도가 한국 고등학교 학생들의
영어 어휘 학습에 미치는 영향

2015년 2월

서울대학교 대학원

외국어교육과 영어교육전공

양 시 은

The Effects of the Involvement Load of Tasks
on English Vocabulary Learning of
Korean High School Students

by

SI-EUN YANG

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 2015

The Effects of the Involvement Load of Tasks
on English Vocabulary Learning of
Korean High School Students

과업 관여도가 한국 고등학교 학생들의
영어 어휘 학습에 미치는 영향

지도교수 오 선 영

이 논문을 교육학 석사 학위논문으로 제출함
2015년 2월

서울대학교 대학원
외국어교육과 영어교육전공
양 시 은

양시은의 석사학위논문을 인준함
2015 년 2월

위 원 장 _____
부위원장 _____
위 원 _____

The Effects of the Involvement Load of Tasks
on English Vocabulary Learning of
Korean High School Students

APPROVED BY THESIS COMMITTEE:

Byungmin Lee, COMMITTEE CHAIR

Jin-Wan Kim

Sun-Young Oh

Abstract

This study attempts to test the involvement load hypothesis, proposed by Laufer and Hulstijn (2001). According to this hypothesis, the involvement load, or the degree of elaboration in processing new words, determines retention of new vocabulary. The hypothesis contends that the involvement load can be analyzed and operationalized in terms of three constructs; *motivation*, *search*, and *evaluation*. If validated, this hypothesis can function as the useful criterion for researchers and educators in designing and evaluating vocabulary tasks. Although several studies examined the validity of the hypothesis, the results are inconsistent; therefore, more research is needed on the involvement load hypothesis. Furthermore, the participants of previous studies were mostly university students, and the various aspects of vocabulary learning beyond the L2 words meaning were not explored in terms of the involvement load hypothesis. Proficiency, one of the most important variables in language learning, was not fully investigated, either.

In this study, 118 (57 advanced and 61 intermediate learners) 3rd grade high school students in Korea completed one of three tasks with the different involvement load, and took immediate and one-week delayed post-tests. The two tests were composed of the passive word learning test (the translation of L2 words into L1) and the active word learning test (providing L2 forms equivalent to the

presented L1 meanings). When it came to overall vocabulary learning, the sum of the passive word learning scores and the active word learning scores, the present study provided near complete support for the involvement load hypothesis, not only for the advanced level group, but also for the intermediate group. Both groups showed better performances in the initial learning and the long-term learning stages when the involvement load of the task was higher if they could manage the task.

Regarding the comparison of passive word learning and active word learning, the results revealed that the involvement load of the tasks influenced those two aspects of vocabulary learning differently. The results of the passive word learning test of both proficiency groups provided near full support for the involvement load hypothesis. As for the active word learning test, however, the involvement load hypothesis failed to predict the long-term learning outcomes of both levels; only the superiority of the writing task was found in the advanced learners. This might suggest that the explanative power of the involvement load hypothesis might be confined to the passive word learning, not encompassing all aspects of complex vocabulary learning. The active word learning might be explained better by other theories, such as the transfer appropriateness processing theory, or might be contingent on the actual processing time of the target words. Results and the implications regarding task-based vocabulary instruction are discussed.

Key Words: Involvement load hypothesis, Vocabulary tasks, Vocabulary learning retention, Passive and active word learning

Student Number: 2013-21372

TABLE OF CONTENTS

Abstract.....	i
CHAPTER 1. INTRODUCTION	1
1.1 The Purpose of the Study	1
1.2 Research Questions	8
1.3 Organization of the Thesis.....	9
Chapter 2. LITERATURE REVIEW	10
2.1 Vocabulary in Second Language Learning.....	10
2.2 Extensive Reading and Task-based Vocabulary Learning	12
2.3 Processing Depth Theory	18
2.4 Involvement Load Hypothesis.....	20
2.5 Indirect Evidence for the Involvement Load Hypothesis.....	23
2.6 Direct Evidence for the Involvement Load Hypothesis	27
Chapter 3. METHODOLOGY	35
3.1 Participants	35
3.2 Instruments	39
3.2.1 Text and Target Words.....	40
3.2.2 Tasks	42
3.3 Assessment	49
3.3.1 Passive Word Learning Test.....	49

3.3.2 Active Word Learning Test	50
3.3.3 Scoring	52
3.4 Procedure	53
3.4.1 First session	53
3.4.2 Second session	55
3.5 Data Analysis	56
Chapter 4. RESULTS AND DISCUSSION	58
4.1. Results	58
4.1.1 The Effect of the Task Involvement Load on the Immediate Vocabulary Learning and its Retention of the Learners at the Different Proficiency Levels	59
4.1.2. The Effect of the Involvement of Load of Tasks on the Active and Passive Word Learning of the Learners at the Different Proficiencies	64
4.2. Discussion.....	71
4.2.1 The Effect of the Task Involvement Load on the Immediate Vocabulary Learning and its Retention of the Learners at the Different Proficiency Levels	71
4.2.2. The Effect of the Task Involvement Load on the Active and Passive Word Learning of the Learners at the Different Proficiencies	78
Chapter 5. CONCLUSION	88
5.1 Major Findings	88
5.2 Implications.....	91

5.3 Limitations and Suggestions.....	93
REFERENCES.....	96
국문초록.....	116

List of Tables

Table 2.1 Comparison and Analysis of Involvement Load of Vocabulary Tasks.....	24
Table 3.1 Descriptive Statistics of 7 Participating Classes.....	36
Table 3.2 Descriptive Statistics of the Subgroups.....	38
Table 3.3 Homogeneity of the Task Groups.....	39
Table 3.4 The Involvement Load of Three Tasks.....	48
Table 3.5 Lexical Production Scoring Protocol (Barcroft, 2002).....	51
Table 4.1 Descriptive Statistics of the Immediate Test and the Delayed Test.....	60
Table 4.2. Effects of Task and Proficiency on the Immediate and the Delayed Tests.....	61
Table 4.3 Post Hoc Analysis on the Immediate and Delayed Tests....	63
Table 4.4 Descriptive Statistics of Passive and Active Word Knowledge.....	65
Table 4.5 Effects of Task and Proficiency on the Passive and Active Word Learning.....	67
Table 4.6 Post Hoc Test on the Immediate Test.....	69
Table 4.7 Post Hoc Test on the Delayed Test.....	70

List of Figures

Figure 3.1 Task 1- Reading with the Gloss and Answering the Comprehension Questions (Index 1)	44
Figure 3.2 Task 2- Filling in the Blank (Index 2)	46
Figure 3.3 Writing Original Sentences (Index 3)	48

List of Appendices

Appendix 1. Pretest.....	106
Appendix 2. Task 1	107
Appendix 3. Task 2	109
Appendix 4. Task 3	111
Appendix 5. Passive Word Learning Test; Immediate.....	113
Appendix 6. Active Word Learning Test; Immediate	114
Appendix 7. Passive Word Learning Test; Delayed	115

CHAPTER 1.

INTRODUCTION

This chapter introduces the research by presenting the motivation and the organization of the study. Section 1.1 discusses the problem and the purpose of the study. Section 1.2 presents the research questions, and the overall organization of the study is outlined in Section 1.3.

1.1 The Purpose of the Study

Vocabulary is one of the most fundamental elements in language, since the most basic information indispensable for comprehending and expressing meaning is conveyed through the words. Researchers and practitioners focus on the learners' vocabulary knowledge, since it has strong correlation with the learner's general language ability (Anderson & Freebody, 1981; Koda, 1989). Many learners also recognize the importance of vocabulary learning and often list the lack of proper L2 word knowledge as the main reason for difficulty in L2 communication.

This important element in L2 learning represents a grave challenge for language learners. Language learners are expected to know 8,000 – 9,000 word families for understanding the authentic written texts, and 5,000 – 7,000 word families for the oral discourses (Hsueh-Chao & Nation, 2000; Nation, 2006; Schmitt, 2008). This level of vocabulary is never easy for language learners to acquire, especially in the EFL context, where exposure to target language input is very limited. It is no wonder that many language learners fail to reach this vocabulary level and feel challenged and frustrated facing the immense task of vocabulary learning.

Practitioners and learners, however, do not receive a lot of support in vocabulary learning. In fact, vocabulary used to be one of the neglected areas in SLA (Second Language Acquisition) research; the major theorizing work was concentrated on other areas, such as syntax and phonology (Zimmerman, 1997; Candlin, 1998). There is no concrete, broad theoretical framework in L2 vocabulary learning, and teachers and textbooks generally fail to provide clear guidelines in vocabulary learning (Schmitt, 2008).

Particularly in Korean high schools, the class instruction is focused mainly on other language skills, reading and listening, which are tested in the CSAT (College Scholastic Ability Test). Vocabulary learning in Korean high schools is usually left to the learners. Even when vocabulary is covered in the classroom, it rarely goes beyond the presentation of new words or different senses

of known-words. Internalization and retention of vocabulary is deemed as the learners' own responsibility. Without proper support, the learners usually turn to rote learning of vocabulary, memorizing from the word list. Since such rote-learning does not build meaningful connection with learners' existent knowledge, the memorized words usually do not last despite the massive time and efforts spent by the language learners.

After the attempts to find a way for the longer retention of new information, a large amount of memory research has demonstrated that new information has a higher possibility of being stored in the long-term memory when processed in a more meaningful and elaborate way (Craik & Lockhart, 1972; Baddeley, 1978; Laufer & Hulstijn, 2001). It is in line with Processing Depth Theory proposed by Craik and Lockhart (1972); the deeper the initial processing of new information is, the longer this new information lasts in memory.

This depth or elaboration in processing new words is also crucial for the effective L2 vocabulary learning. Since the task can induce the learners to notice and process the target words elaborately, task-based vocabulary learning can be more effective than rote-learning or reading only method (Luppescu & Day, 1993; Cho & Krashen, 1994; Knight, 1994; Paribakht & Wesche, 1997). Many studies have tried to find the characteristics of the more effective vocabulary learning tasks and indeed some patterns have emerged from those studies. For example, the tasks which elicited production or negotiation of certain target words generally led

to higher vocabulary learning (Joe, 1998, Ellis & He, 1999; de la Fuente, 2002). The commonalities of the effective vocabulary tasks found in the previous vocabulary research, however, are not enough to provide complete and systematic explanation of L2 vocabulary learning; each experiment was independently designed and conducted. There exist no framework in regard to L2 vocabulary learning and the needs to develop criterion on which task effectiveness can be systematically judged has been raised (Laufer & Huisman, 2001).

Based on the findings of previous research on vocabulary tasks and Processing Depth Theory (Craik & Lockhart, 1972; Craik & Tulving, 1975), Hulstijn and Laufer (2001) proposed ‘the involvement load hypothesis’, which categorized and operationalized the characteristics of effective tasks in terms of three constructs; *need*, *search*, and *evaluation*. Need refers to the necessity for the learners to understand the target words in task completion; it can be absent (index 0), moderate (index 1), or strong (index 2). Need is moderate when it is externally imposed and it is strong when it comes from the learners’ own volitions. Search is whether the learners make attempts to find meanings of new words; it can be either absent (index 0) or present (index 1). Lastly, evaluation refers to making decisions about new words; it is moderate (index 1) when the learners make choices among different words or different senses of a word based on the context as in the gap-fill task. Task has strong evaluation (index 2) when the decisions on the surrounding words are required as in the composition task. Involvement load, the degree of

learners' engagement in processing target words, can be calculated by summing up the indexes assigned to those three components. It can range from 0 (no need, no search, no evaluation) to 5 (strong need, search, strong evaluation). The main point of the hypothesis is that the higher involvement load the task generates, the more deeply the learners process the target words, and the longer the knowledge of new words lasts in the learners' memories.

If validated, this hypothesis can be of great value; it can function as a useful criterion on which researchers and language educators can manipulate and evaluate the effectiveness of vocabulary tasks. It can provide clearer guidelines on effective vocabulary learning; it can help teachers and material designers to make more informed decisions in choosing and adapting vocabulary tasks for particular learners, reducing language learners' burden in vocabulary learning.

A number of studies were conducted to test the validity of the involvement load hypothesis (Hulstijn & Laufer, 2001; Lee, 2006; Keating, 2008; Kim, 2008; Jing & Jiabin, 2009; Kim & Na, 2010; Xiaohui, 2010; Park, 2011; Sung, 2013). There is still not enough research and the results of the existent research are not consistent. While some studies (Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Jing & Jiabin, 2009) provided near complete or perfect evidence for the involvement load hypothesis, the hypothesis could not account for the deviate in other studies (Lee, 2006; Kim & Na, 2010; Xiaohui, 2010; Park, 2011; Sung, 2013). This inconsistency in the research results call for more studies on the involvement

load hypothesis.

Furthermore, the participants of previous studies were mostly university students with few exceptions: the ESL university learners in Kim's (2008) research and the EFL high school learners in Lee's (2006) and Sung's (2013) research. It is EFL high school learners who are more susceptible to the effectiveness of vocabulary tasks; high school learners receive more class instruction which generally incorporates tasks, and in the EFL environment, L2 vocabulary input outside the classroom is usually very limited. This raises the necessity of conducting the experiment on the involvement load hypothesis with EFL high school learners.

Proficiency, one of the most important variables in language learning, was also not fully explored. Only a few studies explored this important factor (Lee, 2006; Kim, 2008; Kim & Na, 2010; Sung 2013), and the results about the proficiency do not converge. In Kim's (2008) study, the proficiency was not significant; both the advanced learners and the intermediate learners showed similar levels of performance when engaged in the tasks with the different involvement loads. Other studies (Lee, 2006; Kim & Na, 2010; Sung 2013) concluded that proficiency was a significant factor; the high level learners did significantly better than the lower level learners or some tasks with the high involvement load was effective only for the advanced learners. Considering the importance of proficiency factor in language learning, particularly in Korean

context where the huge gap in learner proficiency exists, more studies on the proficiency factor from the point view of the involvement load hypothesis will be beneficial to a great extent.

Moreover, except for just one study (Keating, 2008), the previous empirical studies on the involvement load hypothesis looked into only passive word learning, the meaning, in spite of the various aspects of vocabulary knowledge such as meaning, orthography, syntax, and so on. Keating (2008) compared two aspects of vocabulary knowledge, active word learning (i.e., word form) and passive word learning (i.e., word meaning). The results indicated that the involvement load of tasks influenced those two types of vocabulary learning differently; while passive word learning was generally in line with the hypothesis, active word learning in the delayed test could not be explained by the involvement load hypothesis. Therefore, it is necessary to investigate passive word learning and active word learning to find out how the task-induced involvement impacts the different aspects of vocabulary learning.

In this context, the present study tries to test the involvement load hypothesis directly, measuring both active and passive recall of Korean high school students at the different proficiency levels.

1.2 Research Questions

The focus of the present study is to investigate the effects of the involvement load of tasks on the vocabulary learning of Korean high school learners at the different proficiency levels from the following two perspectives. First, this study looks into the impact of three tasks with the different involvement loads on the overall vocabulary learning of Korean high school learners. Here, the overall learning refers to the sum of the passive word test scores and the active word test scores. Secondly, the passive word learning and the active word learning of the learners are compared in terms of the task-induced involvement load. The participants in this study completed one of three tasks with the different involvement load and took the immediate and delayed post-tests, both in passive word learning and active word learning. Every experimental process was meticulously developed and administrated to answer the following research questions.

1. How does the involvement load of tasks affect immediate English vocabulary learning and its retention of Korean high school learners at the different proficiency levels?

2. How does the involvement load of tasks affect active and passive English word learning of Korean high school learners at the different proficiency levels?

1.3 Organization of the Thesis

The present study consists of five chapters. Chapter 1 introduces the motivation of the study and presents the research questions. Chapter 2 presents an overview of the literature review on vocabulary learning and the involvement load hypothesis. In chapter 3, the methodology of the study is described in terms of the participants, the instruments, the assessment, the procedure, and the data analysis. Chapter 4 presents the results and discusses the research findings. Finally, Chapter 5 concludes the research with the summary of the major findings, and presents the implications of the present study and the suggestions for further studies.

Chapter 2.

LITERATURE REVIEW

This chapter presents the literature overview related to L2 vocabulary learning and the involvement load hypothesis. Section 2.1 discusses vocabulary learning in second language. Extensive reading and task-based vocabulary learning are compared in Section 2.2. Section 2.3 deals with Processing Depth Theory, which the involvement load hypothesis, the topic of Section 2.4, was based upon. In Section 2.5, the interpretation of the previous vocabulary task studies from the viewpoint of the involvement load hypotheses is discussed as the indirect evidence. Finally, Section 2.6 presents the empirical studies, which directly tested the involvement load hypothesis.

2.1 Vocabulary in Second Language Learning

Vocabulary has been considered one of the most important factors in second language education. Candlin (1998) went even further to suggest the centrality of vocabulary in language education, saying “the vocabulary is at the heart of language teaching, in terms of organization of syllabuses, the evaluation of learner performance, and the provision of learning resources (p. 1).” Many

learners recognize the importance of vocabulary and report the experiences of difficulties or even breakdowns in L2 communication due to the lack of adequate vocabulary. Vocabulary has been emphasized as the fundamental element in language, which conveys the basic information necessary for comprehending and expressing the message in the target language. Teachers also focus on vocabulary since learner's general language ability, especially reading ability, can be inferred from the vocabulary size, since they usually correspond to each other (Anderson & Freebody, 1981; Koda, 1989).

Vocabulary, the crucial area, however, is never easy to conquer. After reviewing the studies on L2 vocabulary learning, Schmitt (2008) concluded that English language learners should have the vocabulary knowledge of 8,000-9,000 word families for reading, and 5,000-7,000 for speaking and listening. Hsueh-Chao and Nation (2000) calculated the minimum percentage of vocabulary knowledge for language learners to understand written or oral discourse; the result was 98-99%, 49 words out of 50 words. Nation (2006) reached similar conclusion based on the analysis of British National Corpus; language learners have to know about 8,000-9,000 word families to comprehend authentic reading materials and 6,000-7,000 word families for spoken materials. Not surprisingly, many language students fail to reach this level (Nation, 2006). This challenge is especially harder for EFL learners than ESL learners, who receive more limited input and less instruction not only on vocabulary but also on the target language itself.

The learners and the teachers, however, are not sure of the best ways for the vocabulary learning, and the textbooks rarely present clear guidelines on it (Schmitt, 2008). This makes the challenge of vocabulary learning even more troublesome for the learners. This is partly because vocabulary has not received focal attention in the field of SLA; the major theorizing work used to be done in other areas, such as syntax and phonology (Zimmerman, 1997; Candlin, 1998). Nowadays, however, the importance of vocabulary in L2 learning has become much clearer to learners, researchers, and practitioners. As a result, vocabulary has received increasing attention in the SLA field and a large body of research has been carried out on vocabulary learning, which is presented in the following section.

2.2 Extensive Reading and Task-based Vocabulary Learning

In L1 vocabulary learning, extensive reading has been considered as the most fundamental and effective way. After calculating the vocabulary size that children are expected to learn and the amount of words presented in class instruction, Nagy and Anderson (1984) stated that “even the most ruthlessly systematic and direct vocabulary instruction could neither account for a significant proportion of all the words children actually learn, nor cover more than a modest

proportion of the words they will encounter in school reading materials.” (p. 304). The same line of reasoning is applied to L2 vocabulary learning; 8,000 – 9,000 family words, the estimated vocabulary level for efficient comprehension, cannot be covered in the classroom, so the most proportion of vocabulary learning is believed to occur from extensive reading (Pellicer-Sánchez & Schmitt, 2010). Krashen (1989), the strongest proponent of extensive reading for vocabulary learning, argued “competence in spelling and vocabulary is most efficiently attained by comprehensible input in the form of reading” (p. 440). Many other researchers also advocated extensive reading as the ultimate way for second language vocabulary learning (Day, Omura, & Hiramatsu, 1992; Day & Bamford, 1998; Day & Swan, 1998; Tekman & Daloğlu, 2006).

Despite the contribution of extensive reading to L2 vocabulary development, the limitations of extensive reading were also pointed out, especially in the EFL context. The most cited weakness of extensive reading for vocabulary learning is its low pick-up rate (Schmitt, 2008). In Nagy, Anderson, and Herman’s (1987) study on the elementary school students’ L1 vocabulary learning through extensive reading, the students learned only one word out of 20 unfamiliar words. Horst, Cobb, and Meara (1998) reached similar conclusion after reviewing the extensive reading studies in L2; the overall pick-up rate of unfamiliar L2 words in extensive reading was about one in twelve words. According to Hill and Laufer (2003), at this pick-up rate, L2 learners would have to read more than 8 million

words, 420 books, to reach the vocabulary level of 2,000 word families. This is an unrealistic and discouraging standard for L2 readers. Zahar, Cobb, and Spada (2001) also calculated how long it would take for L2 learners to reach 2,000 words through reading if one class hour a week is spent on extensive reading for 40 weeks a year. They reached another daunting answer, 26 years.

In addition to the low pick-up rate, 'Beginners Paradox' (Coady, 1997) is also mentioned as the possible drawback of vocabulary learning through extensive reading; the learners should possess the threshold level vocabulary to learn new vocabulary from reading. This paradox is especially relevant to L2 learners, whose vocabulary knowledge is usually very limited. The learners whose L2 vocabulary level does not reach the threshold level cannot understand the L2 text well, let alone figuring out and retaining the meaning of unfamiliar words.

Another limitation of extensive reading in vocabulary learning is learners' tendency to ignore new words rather than trying to guess the meanings of unfamiliar words from the context and remembering them (Fraser, 1999; Paribakht & Wesche, 1999). In addition, when learners tried to infer the meaning of unknown words from the context, their guesses often were incorrect (Bensoussan & Laufer, 1984). Learners might not be able to incorporate the correct meanings of unfamiliar words into their mental lexicon or would not even attempt to do so if there is no special need to pay attention to new words (Laufer & Shmueli, 1997).

The limitations of extensive reading discussed above led many

researchers and practitioners to a more balanced position, the integration of extensive reading and more explicit vocabulary teaching approach (Tekmen & Daloğlu, 2006; Schmitt, 2008; Pellicer-Sánchez & Schmitt, 2010). After reviewing L2 vocabulary research, Schmitt (2008) concluded that the incorporation of explicit vocabulary approach yielded greater and faster learning, longer retention, and even productive mastery of new words.

Task is often used to lead the learners' attention explicitly to the target vocabulary while engaging learners in other L2 skills, such as reading, speaking, and so on. Even though focus on form was widely discussed in grammar acquisition, it was rarely discussed in L2 vocabulary learning area. Laufer (2005), however, emphasized the importance of the form-focused approach in L2 vocabulary learning, the implementation of various vocabulary-focused tasks. In fact, a significant amount of research provided the evidence that the incorporation of vocabulary task was more effective in vocabulary learning and its retention than reading alone (Luppescu & Day, 1993; Cho & Krashen, 1994; Knight, 1994; Paribakht & Wesche, 1997).

For example, Paribakht and Wesche (1997) randomly assigned 38 young adult beginner and intermediate ESL learners to one of two instruction conditions; reading only and reading plus vocabulary enhancing activities. The subjects' vocabulary gains were measured using 'Vocabulary Knowledge Scale (VKS)'. In this scoring system, the score ranges from unfamiliarity of words through

recognition of meaning, knowledge of meaning, to semantically and grammatically accurate use of the words in a sentence. The research results showed that reading plus vocabulary enhancing activities group learned significantly more vocabulary than the reading only group. Furthermore, while the vocabulary knowledge of the reading only group stayed in the recognition level, the learners in the reading plus vocabulary activities group reached higher knowledge level. Similarly, in Luppescu and Day's (1993) study, Japanese university learners of English who looked up the unknown words in the bilingual dictionary while reading showed 50 % greater vocabulary learning gains than those who read the text without access to the dictionary.

Accepting the merits of vocabulary task, many studies have compared various vocabulary-enhancing tasks, and found that certain types of vocabulary tasks were more effective. The tasks which involve negotiation or production of the target vocabulary engaged learners in more active processing of the target words and yielded better vocabulary learning than the tasks which did not require much mental effort from the participants (Joe, 1998; Ellis & He, 1999; de la Fuente, 2002).

For example, Ellis and He (1999) concluded that learners' receptive and productive vocabulary gains were significantly better when they had chances to use the target words to negotiate the meaning (interactionally modified output) than when they were mainly exposed to the target words (premodified input). De

la Fuente (2002) reached similar conclusion. In receptive vocabulary learning, both the modified input and interactionally modified output group outperformed the premodified input group. In the productive vocabulary test, the interactionally modified output group showed significantly better learning outcomes than the other two groups. Also in Joe's (1998) study, adult English learners in ESL context showed more efficient vocabulary learning when they retold the content of the text without access to the original text than when they had the access to the original text. The target words were used in more learner-generated context in the former condition, while the participants used the target words almost in the same way as in the input text in the latter condition. Joe concluded that the higher generating process of the target words is conducive to more effective productive and receptive vocabulary learning of unfamiliar words.

Although some patterns have emerged in regard to the characteristics of more effective vocabulary tasks, the previous vocabulary task research cannot provide the complete and systematic explanation of vocabulary learning; each study independently analyzed different type of tasks with the diverse standards. This deficiency and the necessity of a more general and systematic framework in L2 vocabulary learning has been addressed by researchers (Laufer & Hutasjin, 2001). In an attempt to provide a framework for analyzing the effectiveness of vocabulary tasks in a more systematic way, Laufer and Hutasjin (2001) proposed "the involvement load hypothesis", based on the review of the previous vocabulary

task research and “Processing Depth Theory” proposed by Craik and Lockhart (1972). The theory claims that the more effective task requires a deeper processing of new words.

2.3 Processing Depth Theory

The “Processing Depth Theory”, proposed by Craik and Lockhart (1972), was responsible for shifting the focus in memory research from the structural aspects of the memory storage to the encoding process. While the previous information model assumes separate memory storages and the different coding processes, processing depth theory “focuses on the encoding operations themselves”, claiming that “the rates of forgetting are a function of the type and the depth of encoding process” (p. 675). This theory recognizes different levels of processing; lower stages refer to handling of rather physical and sensory stimulus while deeper processing involves recognizing patterns, figuring out the meaning of input, comparing the new information with the old information, and so on. These different levels of processing stage are the depths of processing. The more deeply the input is processed initially, the stronger memory traces are built, and the longer new information is retained. Later, Craik and Tulving (1975) modified the theory by specifying the “depth” of processing as “greater semantic involvement” (p. 268). They further developed the theory by claiming that the

information processed in a more compatible context is remembered better and the degree of elaboration and semantic decision is more crucial than the mere depth of processing.

The theory was, however, not free from criticism. The major criticism comes from its insufficient explanation of the processing mechanism and the lack of an independent measure for the depth of processing (Baddeley, 1978). An independent measure of the level of processing is required for the adequate development of the concept. As Baddeley (1978) pointed out, however, Craik and Tulving's (1975) efforts to explore the processing time as a measure of the processing depth failed, and the theory has yet to have the breakthrough. The concept is rather circular; the processing which yields better retention is assumed to involve deeper processing, or if the processing results are poor, it is considered as shallow.

Despite its criticism, it is still a widely accepted concept by many researchers that processing new information more elaborately generally leads to better retention (Laufer & Hulstijn, 2001). Baddeley (1978) wrote that the criticism of the processing depth theory does not prove it is wrong, but requires further modification and clarification. Craik and Lockhart (1972), the founders of the theory, also acknowledged the shortcomings of the theory, such as the lack of the operational definitions of 'depth' and 'elaboration', but they accentuated the heuristic value of the theory as a framework for new research. Efforts to improve

the theory were made by Hulstijn and Laufer (2001), who tried to relate the theory to L2 vocabulary acquisition. By providing more operational definition of the depth of processing in terms of the motivational-cognitive constructs, Hulstijn and Laufer (2001) developed the involvement load hypothesis.

2.4 Involvement Load Hypothesis

Based on the results of the vocabulary task research and processing depth theory, Hulstijn and Laufer (2001) proposed ‘the involvement load hypothesis.’ Involvement load means the degree of learners’ engagement in processing new vocabulary. Hulstijn and Laufer defined ‘involvement load’ as “a motivational-cognitive construct which can explain and predict learners’ success in the retention of unfamiliar words” (p. 14). The hypothesis is based on the cognitive construct, processing depth theory, which was discussed in Section 2.3, and the motivational construct; Hulstijn and Laufer (2001) acknowledged that L2 learners are not information-processing devices and their information processing is affected by motivation and emotion.

The construct of involvement can be analyzed and operationalized by three components, *need*, *search*, and *evaluation*; *need* is a non-cognitive and motivational construct and the latter two are cognitive constructs. Each of those components can be absent or present and have different prominence, e.g. moderate

or strong. The sum of the scores of those factors indicates involvement load. This involvement load can be operationalized and generated by the task; the vocabulary task designers can manipulate the presence or the degree of those three components. The hypothesis contends that the learners process the target words more deeply and retain them better when the vocabulary task has a higher involvement load. In other words, the higher the involvement load of task is, the more effective the task is in retention of new vocabulary.

Need is the non-cognitive, motivational component; whether the understanding of target words is required to complete the task. This component has two degrees of prominence, moderate or strong. It is moderate when externally imposed, and it is strong when imposed by the learners themselves. For example, when the learners are asked to write a sentence using the target word provided by the teacher, need in this task is moderate since the external agent generates it. Need is strong when the learners lack the knowledge of L2 words they need to communicate in L2.

Search, the cognitive construct, can only be present or absent; it does not have different prominence according to Hulstijn and Laufer (2001). Search is present when the learners make attempts to find the meanings of L2 words or the L2 words equivalent to L1 words by looking up the words in a dictionary or resorting to other authorities, such as teachers. If the task allows the learners to use the dictionary and the students do look for the meanings of unfamiliar L2

words, the task has the search component. In contrast, if the task provides a gloss of new L2 words, the search component is absent.

Evaluation is a cognitive component, and has two degrees of prominence; moderate (1) and strong (2). Evaluation refers to “making a selective decision based on a criterion of semantic and formal appropriateness of the word and its context (Laufer & Hulstijn, 2001, p.14).” It is considered ‘moderate’ when the task requires the learners “to recognize differences between words, or differences between several senses of a word in a given context (ibid, p. 14).” It is referred to as ‘strong’, if the task requires “making a decision about additional words which will combine with the new word in an original sentence or text (ibid, p. 14).” For example, when the learners engage in the fill-in task with the words and meanings provided, the task has moderate evaluation since the students have to make decisions by comparing the given words against the context. The composition task generates strong evaluation; the learners have to make decisions about the words which co-occur with the target L2 words.

In short, the task can generate different degrees of involvement load and can be manipulated and analyzed in terms of three components, need, search, and evaluation. The involvement load is indicated by the sum of those three components; the involvement load may range from 0 (no need, no search, no evaluation) to 5 (strong need, search, strong evaluation). The higher the involvement load that the task generates, the more effective the task is in long-

term retention of target words.

2.5 Indirect Evidence for the Involvement Load

Hypothesis

Laufer and Hulstijn (2001) returned to previous vocabulary task studies and checked whether the effectiveness of the tasks in those studies could be explained by the involvement load of the tasks. They found that the tasks, which had not been initially designed for testing the involvement load hypothesis, showed the relationship between the involvement load of the tasks and their effectiveness. Based on the format in Laufer and Hulstijn (2001), Table 2.1 displays the more effective tasks and the less effective tasks in the original studies in terms of the composition of three involvement components (need, search, and evaluation).

As in Table 2.1, the tasks in which the participants looked up unknown words in the dictionary were found to be more effective than the tasks where the students just read the text or were provided with the gloss of new words (Luppescu & Day, 1993; Cho & Krashen, 1994; Knight, 1994; Hulstijn et al., 1996; Laufer, 2000). In Cho and Krashen's (1994) case study of four ESL learners, two learners who were allowed to use the dictionary and consulted it remembered new

Table 2.1

Comparison and Analysis of Involvement Load of Vocabulary Tasks

Study	More effective treatment	Less effective treatment
Lupescu & Day (1993)	reading + dictionary use [4]	Reading only [0] or [2]
Cho & Krashen, (1994) Knight (1994)	: need 2, search 1, evaluation 1	: need (0/1) search (0/1) evaluation (0)
Hulstijn et al. (1996)	Reading + dictionary use [4]	Reading + marginal gloss [1]
Laufer (2000)	: need 2, search 1, evaluation 1	: need 1, search 0, evaluation 0
Ellis et al. (1994)	Interactionally modified input [3] : need 2, search 1, evaluation 0	Premodified input [1] : need 1, search 0, evaluation 0
Ellis & He (1999)	Interactionally modified output [5]	Interactionally modified input [3]
de la Fuente (2002)	: need 2, search 1, evaluation 2	: need 2, search 1, evaluation 0 Premodified input [1] : need 1, search 0, evaluation 0
Hulstijn (1992)	Selecting meaning of words from several choices [2] : need 1, search 0, evaluation 1	Meaning provided in synonym [1] : need 1, search 0, evaluation 0
Joe (1998)	Words used in the original context	Words used in the non-original
Kitajima (2001)	[3]	context or presented as input [1]
Webb (2005)	: need 1, search 0, evaluation 2	: need 1, search 0, evaluation 0
Peters et al. (2009)	Comprehension questions relevant to the target words [2] : need 1, search 1, evaluation 0	Comprehension questions irrelevant to the target words [0] : need 0, search 0, evaluation 0

vocabulary more than the other two subjects to whom the dictionary use was forbidden. The reading with the dictionary task has the involvement load index (4). It generated ‘strong need (2)’ and ‘search (1)’, since the decision and the act to look up the words came from the participants’ volition. It has ‘moderate evaluation (1)’ because the learners had to choose one sense, which fit the context best. Meanwhile, the reading only task has two possible involvement indexes, 0 or 2; moderate need (1) and search (1), if the learners tried to guess the meaning of the unknown words from the context, and no involvement load at all if the learners decided to ignore the new words. Another less effective task, reading with gloss, has an involvement load of 1, since there exists only moderate need (1); no search and evaluation is present.

In terms of the involvement load hypothesis, the learners were found to remember new words better when engaged in negotiation for output than when they interactionally modified the input, which in turn was superior to when they just received pre-modified input (Ellis et al., 1994; Ellis & He, 1999; de la Fuente, 2002). The interactionally modified output tasks (index 5) had strong need (2), search (1), and strong evaluation (2), since the learners used the target word actively in the original contexts for communication. In the negotiation of input task (index 3), there is no evaluation, but just strong need (2) and search (1). When the learners had no opportunities to negotiate the meaning of new words, the involvement load index was 1; moderate need (1), no search (0), and no evaluation

(0).

For instance, in Ellis, Tanaka, and Yamazaki's (1994) study, 206 Japanese EFL high-school students received two different treatments. The pre-modified input group listened to the directions about the specific positions of the kitchen utensils from the native speaker teacher of English and chose the pictures the most suitable for the descriptions, and was not allowed to interact with the teacher. The learners in the interactionally modified input group received the same directions from the native speaking teachers, but they could make request for repetition or clarification of many items. In the post vocabulary test, the learners in the latter group (index 3) noticed and retained unknown words better than the former group (index 1).

Research was also conducted on the tasks which instructed the learners to use the target words in the original contexts (Joe, 1998; Kitajima, 2001; Webb, 2005). For instance, Webb (2005) gave two different treatments to the EFL Japanese university students and administered the 10 types of vocabulary tests measuring the various aspects of target words, such as orthography, meaning, syntax, and so on. The learners who were asked to write one sentence using the target word scored significantly better in all 10 tests than the students who read three glossed sentences. When the learners used the target words in the original context with the provided gloss, the task had the involvement load index 3; the learners wrote the sentence since the task required it (need 1), they did not have to

look up the words (search 0), and the learners had to make plenty of decisions about the co-occurring words (evaluation 2). The tasks in which the learners understood the input or could resort to the original text when retelling generated only moderate need (1), but no search (0), and no evaluation (0).

Even though the previous research results can be explained by the involvement load hypothesis, it should be remembered that it is only indirect evidence, the interpretation from the involvement load hypothesis. The tasks were not designed for the involvement load hypothesis, and there is no guarantee that the research results were not influenced by the variables other than the involvement load of the tasks, such as the topic, or the frequency of the target words in the input. Therefore, more empirical experiments, which directly test the involvement load hypothesis, are necessary to validate the hypothesis and to prove its usefulness (Laufer & Hulstijn, 2001).

2.6 Direct Evidence for the Involvement Load Hypothesis

To date, several empirical studies were conducted to investigate the involvement load hypothesis (Hulstijn & Laufer, 2001; Lee, 2006; Keating, 2008; Kim, 2008; Jing & Jiabin, 2009; Kim & Na, 2010; Xiaohui, 2010; Park, 2011; Sung, 2013). The results of those studies, however, are not conclusive yet. This gives rise to the necessity of more empirical studies on the involvement load

hypothesis. The results of the Hebrew group in Hulstijn and Laufer (2001)'s study, the ESL young adults in Kim (2008)'s research, and the EFL Chinese students in Jing and Jiabin's (2009) research provided near full support for the involvement load hypothesis; the retention of newly learned words were in line with the prediction of the involvement load hypothesis. For example, the Israeli university students in Hulstijn and Laufer (2001)'s research showed the best performance when they completed the composition using the target words (index 3). The learners who completed the gap-fill task (index 2) scored the second best, and the learners did the least successfully when the task was answering comprehension questions with the marginal gloss (index 1).

The Dutch group in Laufer and Hulstijn (2001)'s study, however, provided partial evidence; while the composition group (index 3) and the gap-fill group (index 2) yielded results consistent with the prediction of the hypothesis, the gap-fill group did not do significantly better than the reading with the marginal gloss group (index 1). Similar partial results were found in Keating (2008)'s research with the 79 beginner learners of Spanish; the writing original sentence (index 3) and the gap-fill task (index 2) were superior to the reading only task (index 1) in learning pseudo words, but the sentence writing task (index 3) was not significantly better than the gap-fill task (index 2).

On the other hand, some studies have produced results significantly deviate from the prediction of the involvement load hypothesis (Lee, 2006; Kim

& Na, 2010; Xiaohui, 2010; Park, 2011; Sung, 2013). For example, in Kim and Na (2010)'s research, the Korean University learners' initial learning, which was consistent with the involvement load hypothesis, disappeared in the 4-week delayed test; only the composition group (index 3) was significantly better than the reading only group (index 1), and there were no further significant differences.

After Hultijin and Laufer (2001)'s initial success to provide the complete and partial support for the involvement load hypothesis, a few studies moved on to testing the hypothesis with the different tasks. In many cases, however, the interpretation of the alternative tasks was rather arbitrary, and the variables other than the task-induced involvement load were not meticulously controlled. This might have been the cause of the unsuccessful results of those studies mentioned before. For example, in the experiment with the Korean high school learners, Sung (2013) used the unscrambling sentence task as the alternative for the writing original sentence (index 3); the learners were instructed to write English sentences equivalent to the presented Korean sentences by reordering the given English words. This decision was made out of the worries that Korean high school learners would not be able to write the original sentences. Even though the learner-generated context was crucial for strong evaluation, the researcher regarded the fore-mentioned two tasks as having the same involvement load. The research results that the unscrambling task (index 3) was not significantly better than the gap-fill task (index 2) might have resulted from this arbitrary interpretation of the

new task.

It is also worth noting that the tasks which implemented the dictionary use generally resulted into the unsatisfactory results for the hypothesis (Xiaohui, 2010; Park, 2011). The dictionary use task was intended to generate search and moderate evaluation; the learners were supposed to look up the words and to find the right sense for the context. There was still no guarantee, however, that the learners would not ignore the unfamiliar words, or would not end up with the wrong meaning. In those cases, the intended involvement load was lost due to the failure of the variable control.

For example, Hulstijn et al. (1996) found that the participants in the dictionary use group remembered the target words to a greater extent than the reading only group when the dictionary group learners actually referred to the dictionary. But this superiority of the dictionary group disappeared when the learners in the dictionary group decided not to look up the words. Lee's (2006) result also suggests the importance of the variable control in the dictionary use task. In her research, the learners in the gap fill with the dictionary task (index 3) were instructed to look up the target words and to write down the first meaning of each entry to ensure that the learners in the dictionary group did search the words and found the right meanings. The vocabulary test results of the well-controlled gap-fill with the dictionary task (index 3) turned out to be as effective as the sentence writing task (index 3) as hypothesized by the researcher. Even though the

use of the new tasks in testing the hypothesis is worth praises, the problems discussed above might indicate the appropriateness of using the more basic tasks from the previous research fully, before moving on to the innovative new tasks.

In addition, the previous empirical studies on the involvement load hypothesis did not explore the learner variety completely. Most of the studies were conducted on the EFL university learners with very few exceptions; ESL university learners in Kim's (2008) study, and EFL high school learners in Lee's (2006) research and Sung's (2013) research. In fact, the high school learners receive more class instruction, which often implements tasks, and task effectiveness is especially important in the EFL context, where the exposure to the target language and the learning opportunities are quite restricted. Therefore, the involvement load studies with EFL high school learners can be very beneficial.

Furthermore, proficiency, one of the most important learner variables, was explored only in a small number of studies (Lee, 2006; Kim, 2008; Kim & Na, 2010; Sung 2013). Laufer and Hulstijn (2001) claimed that the learners at the different proficiency levels can benefit from the task with the high involvement load if they can manage the task. Moreover, the results about the proficiency in those studies were not consistent. Kim (2008) concluded that the proficiency was not a significant factor; the advanced level learners and the intermediate level learners were almost equally successful in retention of the new words in the tasks with the different involvement load. On the contrary, the other three studies

revealed that the learner proficiency was a significant factor; the more proficient learners performed significantly better than the less advanced learners, or some tasks with higher involvement load were effective only for the advanced learners.

The necessity to explore the proficiency further also lies in the inadequate labeling of learner proficiency in the previous research. For example, Kim's (2008) research was criticized since the undergraduate ESL learners with TOFEL score above 520, labeled as the intermediate group in her study, should have been classified as the advanced learners. Both groups in Kim's study might have belonged to the advanced group, and this might have been the reason for the insignificance of the proficiency factor in her study. Other indelicate proficiency classifications were found in Sung (2013)'s and Lee's (2006) research. In the former research, the learners within the top 21 % in the nation-wide practice CAST English test were classified as the advanced level, and the rest of the learners (22 – 100%) as the intermediate. Lee divided the learners based on in-school English test score, which has less credibility than the CAST test; the learners with above 50 of 100 points belonged to the advanced group and the students below 50 points to the intermediate group. Such inappropriate proficiency classification in past studies indicates that it is necessary to study the proficiency factor with more rigorous proficiency labeling.

Another gap in the previous research is that the investigation was mostly focused on passive word learning, the meaning, despite the complex aspects of vocabulary knowledge. To the researcher's knowledge, only one study compared active word learning (i.e., word form) and passive word learning (i.e., word meaning) gains. According to Keating's (2008) results, while the passive word test score provided partial support for the hypothesis as described earlier, active word learning did not hold in the delayed test two weeks later. In the immediate test, the active word test scores were in order of the writing sentence task (index 3), the gap-fill task (index 2), and the reading and answering comprehension questions (index 1) in accordance with the involvement load hypothesis. In the delayed test, however, the gap-fill task was significantly better than the reading task, but the writing task was not better than either of the former two tasks. To consolidate or refute Keating's research findings that the involvement load of tasks influences the active and passive word learning differently, more studies should look into those two aspects of vocabulary learning in terms of the task-induced involvement load.

As discussed so far, the inconsistency of the research results and the fact that such diverse results might have been attributed to the lack of thorough variable control or the misinterpretation of the involvement load of tasks call for more empirical research with more thoroughly designed tasks. The research on high school learners in the EFL context, the least explored learners, can be especially

beneficial. The EFL context is limited in terms of the learning opportunities and the high school learners receive more class instruction than the university learners, thus more susceptible to the task effectiveness. In addition, the proficiency was not fully studied and the existing results do not converge, and this requires more studies with more careful proficiency classification. In addition, the empirical study on more than one aspects of the word knowledge is necessary since the most of the previous research is confined to the semantic area in spite of the complex and various aspects of vocabulary knowledge.

Based on the findings and the limitations found in the empirical research on the involvement load hypothesis, the present study plans to look into the influence of the involvement load of tasks on the immediate vocabulary learning and its retention of Korean high school learners at the different proficiency levels. The learning gains will be measured in the active word learning as well as in the passive word learning for more comprehensive understanding of the impact of the involvement load on L2 vocabulary learning.

Chapter 3.

METHODOLOGY

This chapter describes the methodology employed in the present study. Section 3.1 discusses the participants. Section 3.2 provides details on the instruments in terms of the text, the target words, and the tasks. The word learning assessment methods and their scoring procedures are described in Section 3.3. The procedure and the data analysis are explained in Sections 3.4 and 3.5, respectively.

3.1 Participants

This study was conducted from July to September, 2014. All the participants (110 girls and 148 boys) were third grade high school learners from one co-educational S high school, located in Mok-dong, Yangcheon Gu in Seoul, district well known for its educational fervor. The school has a higher percentage of high and intermediate levels English learners than the rest of the Korean high schools. According to the Ministry of Education, the school ranked 107th out of 2,282 high schools in Korea in terms of the number of students who scored within the band 2 (the top 11%) in 2013 CSAT (College Scholastic Ability Test). Most of

the learners have had at least 9 years of English education, three years in the elementary school, three in the middle school, and three in the high school.

Seven intact classes were chosen based on the mean scores of English section in the nation-wide practice CSAT, which had been administrated on March 12th, 2014. Table 3.1 shows the mean scores and the standard deviations of the CSAT English scores of 7 participating classes. One-way analysis of variance (ANOVA) confirmed the homogeneity of the participating classes ($p = .434$). One of three tasks with the different involvement load index was randomly assigned to those 7 homogeneous classes; task 1 to the classes D and E (72 students), task 2 to the classes A, B, C (112 students), and task 3 to the classes F and G (74 learners).

Table 3.1

Descriptive Statistics of 7 Participating Classes

	Class(Number)							
	A(41)	B(35)	C(36)	D(36)	E(36)	F(35)	G(39)	Total(258)
<i>M</i>	114.37	116.40	118.92	117.44	121.11	115.71	116.64	117.17
<i>S D</i>	17.39	20.97	19.81	26.75	14.97	20.53	21.22	20.33

Note. The maximum test score was 143.

From the collected data, two subsets were chosen for the analysis based

on the proficiency: top 1-10% for the high level, and top 20-50% for the intermediate level. The learners in the 11-19% range were left out of the study to ensure clear distinction between two proficiency groups. This decision was made based on Kim's (2008) research, in which both proficiency groups were not clearly distinguished; both levels were actually high-level learners. This unclear level distinction might have been the reason for the insignificance of the proficiency factor as admitted by Kim herself and criticized by Kim and Na (2010). In the present study, the two proficiency groups were significantly different from each other ($p=0.000$). In addition, the learners ranked below 51 % were also excluded from the further analysis due to the small sample size (5 for task 1, 11 for task 2, and 7 for task 3). The small sample size resulted from the high percentage of advanced and intermediate learners of the target school, and also from the high absence rate of the low level learners; 18 learners out of 41 pertaining to the low level missed the immediate post-test.

To make the group truly representative of the students in the EFL context, all participants with more than 6-months long experiences abroad (61 learners) were ruled out. The learners who identified more than 4 items as pre-known words or as re-encountered words between the immediate test and the delayed test were also excluded from the data analysis. As a result, 57 high level learners (18 for task 1, 19 for task 2, 20 for task 3) and 61 intermediate level learners (21 for task 1, 21 for task 2, 19 for task 3) were selected for the data analysis. Table 3.2

describes the mean scores and the standard deviations of the task groups.

Table 3.2

Descriptive Statistics of the Subgroups

	Whole Group (number)			Advanced (number)			Intermediate (number)		
	Task1	Task2	Task3	Task1	Task2	Task3	Task1	Task2	Task3
	(39)	(40)	(39)	(18)	(19)	(20)	(21)	(21)	(19)
<i>M</i>	122.93	121.18	123.08	137.68	136.56	137.80	109.57	108.00	107.58
<i>SD</i>	2.46	2.50	2.50	4.52	3.74	4.34	5.28	6.89	6.48

Note. The maximum test score was 143.

The homogeneity among the task groups was checked through one way ANOVA with the task type set as the independent variable and the English CSAT standard scores set as the dependent variable. As seen in Table 3.3, task groups in all proficiencies were homogeneous in terms of the practice English CSAT scores; *p* values were 0.836, 0.616, and 0.565 for the whole group, the advanced group, and the intermediate group. On the other hand, two proficiency groups of each task were significantly different from each other (all *ps* = 0.000).

Table 3.3***Homogeneity of the Task Groups***

Whole Group						
	<i>SS</i>	<i>Df</i>	<i>MS</i>	<i>F</i>	<i>P</i>	η^2
Task	86.99	2	43.50	0.18	0.836	0.003
Error	27917.29	115	242.76			
Advanced						
	<i>SS</i>	<i>Df</i>	<i>MS</i>	<i>F</i>	<i>P</i>	η^2
Task	17.51	2	8.78	0.49	0.616	0.018
Error	965.750	54	17.90			
Intermediate						
	<i>SS</i>	<i>Df</i>	<i>MS</i>	<i>F</i>	<i>P</i>	η^2
Task	44.98	2	22.49	0.58	0.565	0.019
Error	2265.77	58	39.07			

* $p < .05$ **3.2 Instruments**

In this section, the selection and adaption of the text and the target

vocabulary are described. The description of the tasks employed in the present study will follow.

3.2.1 Text and Target Words

For the selection of the text used in task 1 and task 2, several ESL reading textbooks were reviewed since they are more readable for the EFL learners than the authentic materials. One text was chosen from *Northstar Reading and Writing, Advanced* (Cohen & Miller, 2004). None of the target participants reported they had studied with this textbook before. The text choice was based on two criteria; 1) the topic is not too specific so that there is little impact of the background knowledge on the text comprehension (Lee, 2007; Nsassaji, 2002), and 2) the percentage of the unfamiliar target words are less than 5%, since the learners should have 95% lexical knowledge to expect by-product learning from reading comprehension (Laufer, 1989; Hsueh-Chao & Nation, 2000; Huang, Willson, & Eslami, 2012). One text titled ‘the Struggle to Be an All-American Girl’ was chosen; the topic, the culture and identity struggles of a second-generation immigrant teenager in the United States, was assumed to be general. The text consisted of 554 words, long enough to generate 10 target words.

Following previous research (Kim, 2008; Kim & Na, 2010; Sung, 2013), which used 10 new words, 10 target words were chosen in the present study. First,

a total of 25 candidate target words (5 verbs, 11 adjectives, 9 nouns) were selected by the researcher and were consulted with the teacher, who had taught the participants for two and half years. Seven words, which were recently learned or presumed to be pre-known by the target participants, were eliminated. Next, with the 18 words (6 verbs, 6 adjectives, 6 nouns), the pretest (see Appendix 1) was conducted on 20 high level students (within the top 10 % in the practice CSAT) of a different school. 10 words, for which none of the pre-test takers provided the right meaning, were selected as the target vocabulary: 3 verbs (*cluck, flank, infuriate*), 3 adjectives (*lilting, musty, raunchy*), and 4 nouns (*blotch, gibberish, ideography, smattering*).

The original text, written for the advanced ESL learners, contained 554 words and a number of complex or long sentences. Considering that the longest text the target learners encounter in the CAST is around 300 words, the text was shorted and modified by the researcher and the current teacher, to prevent the learners from giving up out of frustration. Beside the target words, any words too culturally loaded or not listed as a key word in the national curriculum were replaced with the synonyms. The sentences with complex or unfamiliar structures or with more than 20 words were paraphrased.

The final version, which underwent the review by a native speaker of English, had 378 words. The format was similar to the texts in the CSAT. This final version was tried on the 20 intermediate level learners in the different school

to ensure the readability of the text for the less proficient participants. Those learners reported they had no difficulty comprehending the text and answering the questions.

3.2.2 Tasks

Based on the involvement load hypothesis proposed by Hulstijn and Laufer (2001), three tasks with different involvement load were adopted; reading the text with the marginal gloss and answering the comprehension questions (index 1), filling in the blanks (index 2), and writing original sentences (index 3). In all three tasks, two components, need and search were held constant; 1 for need and 0 for search. The tasks differed only in evaluation; 0 for task 1, 1 for task 2, and 2 for task 3.

3.2.2.1 Task 1

The participants in task 1 group were asked to read the text and to answer four comprehension questions which required the understanding of the target words (see Figure 3.1 for the abridged version or Appendix 2 for the full version). 10 target words in the text were highlighted in bold print and the meaning of target words and the information about the part of speech were provided only in L1

(Korean) in the bottom of the text. The gloss was provided only in L1 for two reasons. First, this is the way the meanings of the new words are presented in the CSAT. Second, L1 gloss can lessen additional cognitive burden in learning new words. Laufer and Shmueli (1997) argued that L1 gloss is more conducive to retention of newly learned words than L2 gloss since learners can pay maximum attention to the target words due to the familiarity and brevity of L1 equivalent. In a similar vein, the task direction was also provided in Korean and the format and the outline of the task was assimilated to the CSAT format. Following the CSAT format, where the longest text usually has 3-4 multiple-choice questions with five choices, task 1 had four multiple-choice questions with 5 choices.

Task 1 had moderate need (1) to refer to the gloss, since it was enforced by the task rather than the learners themselves, no search (0), since the information about the target words was already provided, and no evaluation (0), because the learners did not have to evaluate the target words. Therefore, task 1 had the involvement index 1.

Figure 3.1

Task 1- Reading with the Gloss and Answering the Comprehension Questions

(Index 1)

[1-4] 다음 글을 읽고, 물음에 답하십시오.

It's still there, the Chinese school where my brother and I used to go.
Every day at 5 P.M., my brother and I had to go to Chinese school. My mother was so determined to have us learn Chinese, the language of our heritage. Kicking, screaming, or crying did not stop my mother.
The **musty** classroom smelled like Chinese medicine. I preferred fresh perfume scents of my American teachers in public school. There was a stage, **flanked** by an American flag, and a Chinese flag. I thought the American flag was prettier.

1) **musty**(형용사): 곰팡내 나는, 케케묵은 2) **flank**(동사): ~의 측면에 위치하다

1. 위 글의 제목으로 가장 적절한 것은? [3점]

- ① The Linguistic Differences between Chinese and English
- ② How to Overcome Generation Gaps
- ③ The Importance of English as International Language
- ④ The Struggle to Be an All-American Girl
- ⑤ The Benefits of being Multi-cultural

3.2.2.2 Task 2

Task 2 was a gap-filling activity (see Figure 3.2 for the abridged version or Appendix 3 for the full version). The learners were provided with the same text as in task 1, but 10 target words were replaced by the numbered blanks. The

meaning and the information about the part of speech were presented only in L1 (Korean) in a separate sheet. The information on the words in task 1 and task 2 were exactly the same for the input control. In addition, in task 2, beside the 10 target words, 5 distracters were added to prevent the participants from relying on guessing strategies; if there were 10 words and 10 blanks, the learners would not have to make a careful evaluation for the last one or two words. While in task 1, the target vocabulary was in order of the occurrence in the text, the total 15 words (10 target words and 5 distracter words) were presented in an alphabetical order in task 2.

The participants were instructed to fill in the blanks from the provided word list. Unlike in the previous studies, in which the gap-fill tasks had the same amount of comprehension questions as in the comprehension question tasks (Kim, 2008; Kim & Na, 2010), task 2 of the current study had only two questions, one for the main idea and one for the detail. The inclusion of those two questions was for ensuring that the participants read the text for meaning, not being absorbed only in filling the gaps without understanding the text. This decision was made because, if task 2 includes task 1 as in the previous research, the better results in task 2 would be attributed to the fact that the participants in task 2 also complete task 1; it would be impossible to conclude with certainty that the differences between the task groups come from the different involvement load of the tasks.

Figure 3.2

Task 2- Filling in the Blank (Index 2)

[1-3] 다음 글을 읽고, 물음에 답하십시오.

It's still there, the Chinese school where my brother and I used to go. Every day at 5 P.M., my brother and I had to go to Chinese school. My mother was so determined to have us learn Chinese, the language of our heritage. Kicking, screaming, or crying did not stop my mother.⁴

The 1) _____ classroom smelled like Chinese medicine. I preferred fresh perfume scents of my American teachers in regular school. There was a stage, 2) _____ by an American flag, and a Chinese flag. I thought the American flag was prettier.⁴

1. 아래의 단어 목록 중 적절한 단어를 선택하여 빈칸을 채워 본문을 완성하십시오. [빈칸 당 1점]

- * 선택된 단어들을 단 한 번씩만 사용하십시오.
- * 하나의 빈 칸에는 한 단어만 들어갑니다.

가) abate (동사) 줄다, 감소시키다	다) infuriate (동사) 극도로 화나게 만들다
나) apprehensive (형용사) 염려하는, 두려워하는	라) ideograph (명사) 표의 문자

2. 위 글의 제목으로 가장 적절한 것은? [3점]

- ① The Linguistic Differences between Chinese and English
- ② How to Overcome Generation Gaps
- ③ The Importance of English as International Language
- ④ The Struggle to Be an All-American Girl
- ⑤ The Benefits of Being Multi-cultural

Task 2 had moderate need (1) in that the necessity of understanding the words was imposed by the task, no search (0) since information was already presented, and moderate evaluation (1) because the words in the list had to be compared against the context in the text. The involvement index of task 2 was 2.

3.2.2.3 Task 3

In task 3, the students were required to write 10 original sentences, each sentence including one target word (see Figure 3.3 for the abridged version or Appendix 4 for the full version). A minimum of 7 words for each sentence was recommended. The students were encouraged not to focus too much on the grammar or accuracy but to try their best to write original sentences. The meanings and the parts of speech of 10 target words were provided in alphabetical order in L1. Since Korean learners are not familiar with sentence writing and feel it to be the most burdensome completing these types of productive tasks (Kwon, 2004), an example sentence for each target word was also provided for assistance. These example sentences were necessary since some students, especially the intermediate group learners, might not be competent enough to produce sentences without some guidelines. Those guiding sentences were extracted and modified from the main text of the study for input control; the participants in three task groups received the same information on the target vocabulary.

Task 3 had an involvement index of 3; moderate need (1) since the need to learn the new words was generated by the task, not the learners, no search (0) in that the learners did not have to look up the dictionary or resort to teacher, and strong evaluation (2) because the learners had to make decisions on the surrounding words.

Figure 3.3

Writing Original Sentences (Index 3)

※ 아래의 10개의 단어들을 이용하여 한 문장씩 만들어 보려고 합니다.
 주어진 예시문을 참고하되, 본인의 문장을 영작해 주세요.
 문법이나 철자의 정확성은 중요하지 않습니다. 문장의 뜻을 알 수 있으면 충분합
 니다. (문법이나 철자에 너무 신경 쓰지 마세요)
 각 문장은 가능하면 7개 이상의 단어로 구성해주세요.

1) blotch(명사): 얼룩

예) The pen had to be held in the right way to avoid blotches.

→

Table 3.4

The Involvement Load of Three Tasks

Task type	Involvement Component			Involvement Load Index
	Need	Search	Evaluation	
Task 1 (Answering Questions)	+	-	-	1
Task 2 (Gap-filling)	+	-	+	2
Task 3 (Writing sentences)	+	-	++	3

The involvement load of three tasks in this study is summarized in Table 3.4.

3.3 Assessment

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

3.3.1 Passive Word Learning Test

For the passive word learning test, the learners translated the target words from English to Korean (see Appendix 5). Scoring was conducted by two independent raters; the researcher of the present study, with 3 years of English teaching experience in high school and the other rater, with 10 years of experience of teaching English in middle school. 1 point was awarded for each correct answer and, based on the previous research (Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008), a half point was given to semantically acceptable answers (e.g. answering 'smattering' as 'the lack of knowledge' instead of 'a small number or amount of

something'). In addition, the learners were instructed to pay attention to the part of speech and to reflect it in translation in the test stage. The words which were correct in meaning but incorrect in the part of speech received half a point. The words reported as pre-known or reviewed during the test intervals were scored 0 regardless of whether the test takers provided right or wrong answers. Any confusions and disagreements were resolved through discussions between the raters.

3.3.2 Active Word Learning Test

For the active word learning test (see Appendix 6), the learners were given the list of 10 target words in Korean and were instructed to provide the equivalent English words. 'A lexical production scoring protocol' was adopted for scoring (Barcroft, 2002). This scoring system was chosen because it allows more detailed and precise analysis of word knowledge development in a systematic way. It recognizes not only the complete word form but also the developing partial form. Since the present study measured the initial learning after only one exposure to the target words, the scoring system sensitive to the slight differences in the learners' initial learning was deemed more appropriate. Moreover, since the learners in task 1 (answering the comprehension questions)

Table 3.5

Lexical Production Scoring Protocol (Barcroft, 2002)

Points	Description
.00 points	None of word is written ; this includes <ul style="list-style-type: none">·Nothing is written.·The letters present do not meet any ‘for .25’ criteria.
0.25 points	1/4 word is written ; this includes <ul style="list-style-type: none">· Any 1 letter is correct.· Correct number of syllables is less than 25%.·25-49.9 % of the letters are present
.50 points	1/2 word is written ; this includes <ul style="list-style-type: none">·25-49.9% of letters are correct.·50-74.9 % of the letters are present
.75 points	3/4 word is written ; this includes <ul style="list-style-type: none">·50-99.9 % of letters are correct.·75-99.9 % of the letters are present.·100% of letters correct but other letters are added.
1 point	Entire word is written; this includes <ul style="list-style-type: none">·100% letters correct

were not even provided with chances to write the target words, the scoring system which detects even the slightest change was assumed to be fair to all participants.

In Barcroft's (2002) 'lexical production scoring protocol', scores are awarded based on the percentage of the 'correct' letters, letters in the correct positions, and the percentage of the 'present' letters, letters placed in the incorrect places. For instance, for the target word 'smattering', the answers 'smatter...', '...mattering' were awarded 0.75 points since more than 50 percentage of letters were in their correct positions (70% and 90 %, respectively). The following responses were given 0.25 point; 'st...', 'tam...' since only one letter is correct in the former, and three letters, 30 %, are present in the latter. It was ensured that the raters were trained on this scoring protocol and had plenty of practices before embarking on the real scoring. More detailed scoring system of Barcroft's (2002) lexical production scoring protocol is described in Table 3.5.

3.3.3 Scoring

Two independent experts, who were previously described, scored the test results. Pearson's r was calculated to check inter-rater reliability. The attained values were 0.983 for the passive word learning test and 0.976 for the active word learning test. Due to the very high inter-rater reliability, one of the rater's scores was randomly selected and included in the data analysis. Any items reported as

pre-knowledge by the learners in the immediate test stage or as reviewed during the interval between the immediate test and one-week-later delayed test were excluded from the data analysis.

3.4 Procedure

The procedure of this study had two major sessions. In the first session, 258 high school English learners completed a randomly assigned task and took the post-test, which had three parts; the active word learning test, the questionnaire, and the passive word learning test. In the second session, which took place one week after the first session, the delayed post-test, also composed of two types of vocabulary tests and the questionnaire, was administered. The researcher was present during the entire process. All the tasks and the post-tests were completed during the 50-minute-long class hour with the aid of the current teacher of the participants.

3.4.1 First session

In the first session, 7 homogeneous participant classes were assigned one of three tasks randomly. The learners' written informed consents were obtained

after the general introduction was provided in Korean by the researcher. All the specific directions on task completion were written on the task sheet in Korean for comprehensibility. To prevent rote learning of the target words, two measures were taken based on the operational definitions of the incidental vocabulary learning. First, the learners were not forewarned about the upcoming retention tests (Eysenck, 1982). Second, the tasks were not introduced as vocabulary tasks, following Hulstijn's (2001) definition; incidental word learning happens "as the by-product of any activity not explicitly geared to vocabulary learning (p. 271)." Task 1 and 2 were presented as the reading exercises and task 3 was introduced as the writing practice.

The time spent on each task was not controlled based on Hulstijn and Laufer's (2001) argument that time is an inherent variable of a task. Huang et al. (2012) also pointed out the difficulty in defining "time on task," since the time spent on the task might not exactly correspond to the time the learners actually process the target words. Even though there was prior notification about the time limit, 20 minutes, the learners were instructed to work on their own pace and to record the time they spent on the task. The average task completion time of the advanced level learners was 4.68 minutes, 8.54 minutes, and 12.60 minutes for task 1, 2, 3, respectively. For the intermediate learners, task 1 took 6.52 minutes, task 2, 11.43 minutes, and task 3, 14.26 minutes on average.

Upon completion of the task by all class members, the task sheets were

collected and the immediate post-test, which the learners were not forewarned of, was administrated. The post-test was made up of three parts; the active word learning test (translation the target words into English), the questionnaire, and the passive word learning test (translating the English words into Korean). Three minutes were allowed for each test and there was no time limit on the questionnaire. To prevent the influences of the prior test on the subsequent test, a few measures were taken; 1) the precedence of the active test (Webb, 2005; Webb, 2007), 2) the use of the questionnaire as the distracter between two tests, 3) no pre-warning of the subsequent passive word test in the active word test stage, 4) the different item presentation order in two tests.

3.4.2 Second session

The delayed post-test was conducted one week after the first session. The procedure was the same as in the immediate post-test, but the order of the items on the papers were randomized again to reduce the memory traces. At the end of the passive word learning test in the delayed test (see Appendix 7), the learners were told to indicate any words they reviewed between the immediate post-test and the delayed post-test.

3.5 Data Analysis

From the collected data, the post-test scores of the learners who had more than 6 month experiences in English speaking countries or who marked more than 4 words as pre-known or reviewed were ruled out. The learners whose practice CAST English score was within the top 10 % were classified as the advanced group and the participants who scored between 20% and 50 % as the intermediate group. The data of the advanced level (57), the intermediate level (61), and the combination of the advanced level and the intermediate level (118) were analyzed.

For the first research question, the impact of the level of the involvement load of the tasks on the overall short-term and long-term vocabulary learning, a set of two way ANOVAs (one for the immediate post-test and the other for the delayed post-test) were conducted with task (3 levels) and proficiency (2 levels) set as two independent variables. The sum of the passive word learning test score and the active word learning test score was the dependent variable. Scheffe was used for the *post hoc* analysis.

To answer the second research question, whether the involvement load of tasks influences the passive word learning and the active word learning differently, the data were submitted to a set of two-way MANOVAs. The task types (3 levels) and the proficiency (2 levels) were treated as two independent variables and the passive word learning test score and the active word learning score were set as two

dependent variables. MANOVA was chosen since there was considerable correlation between the active word learning scores and the passive word learning scores. For the whole group, the calculated Pearson's r between the passive word learning test score and the active word learning test score was 0.772 for the short term learning, 0.638 for the long term learning. For the advanced level, it was 0.714 and 0.521, and the intermediate learners' test scores showed the r values of 0.751 and 0.620.

All analyses were conducted using the IBM Statistical Package for the Social Sciences (SPSS) version 21. All the data set met the conditions for using the parametric analysis and the alpha level was set at 0.5.

Chapter 4.

RESULTS AND DISCUSSION

This chapter describes the results of the statistical analysis of the test scores and discusses the findings. Section 4.1 reports the results of the immediate post-test and the delayed post-test around the research questions. Section 4.2 discusses the results.

4.1. Results

Section 4.1.1 discusses the first research question, the effects of the involvement load of tasks on the overall short-term and long-term vocabulary learning gains. Section 4.1.2 reports the results of the second research question, the investigation of the differences between passive and active word learning. In both sections, the results are analyzed in three groups: the advanced, the intermediate, and the whole group (the former two groups combined).

4.1.1 The Effect of the Task Involvement Load on the Immediate Vocabulary Learning and its Retention of the Learners at the Different Proficiency Levels

To investigate the impact of the task involvement load on the general learning gains both in the short term and the long term, the participants (57 advanced learners and 61 intermediate learners) completed one of three tasks with the different involvement index and took the immediate post-test and the delayed test. To analyze the effects of the task on the overall vocabulary learning, the sum of the passive word test score (translation from English to Korean) and the active word test score (translation from Korean to English) was analyzed.

Table 4.1 summarizes the overall means and the standard deviations of the test scores according to three groupings, the combination of two groups, the advanced group, and the intermediate group. The results show that the higher the involvement load of the task was, the better the learners remembered the target words both in the immediate and the delayed tests regardless of their proficiency levels. In all cases, the means of the test results were the highest in the sentence writing task (involvement index 3), the second highest in the gap filling task (involvement index 2), and the lowest in the answering questions task (involvement index 1). For example, the mean scores the advanced group in the immediate test were 5.38, 8.28, and 11.96 for the answering questions task, the.

Table 4.1***Descriptive Statistics of the Immediate Test and the Delayed Test***

Proficiency Level	Task Type	Immediate Test			Delayed Test		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Whole Group (A+ I)	Answering Questions (1)	40	3.67	2.61	39	2.48	2.04
	Gap-filling (2)	39	6.76	3.50	39	3.81	2.78
	Sentence Writing (3)	39	10.12	3.57	39	6.15	2.50
	Total	118	6.82	4.17	117	4.15	2.87
Advanced	Answering Questions (1)	19	5.38	2.40	19	3.74	1.95
	Gap-filling (2)	18	8.28	2.74	18	4.96	2.46
	Sentence Writing (3)	20	11.96	3.20	20	7.84	2.07
	Total	57	8.61	3.90	57	5.56	2.76
Intermediate	Answering Questions (1)	21	2.12	1.67	20	1.28	1.26
	Gap-filling (2)	21	5.45	3.62	21	2.82	2.70
	Sentence Writing (3)	19	8.18	2.89	19	4.38	1.48
	Total	61	5.16	3.74	60	2.80	2.87

Note. The scores are the sum of productive test scores and passive test scores; the maximum score is 20. The numbers in the parentheses indicate the involvement load index of the task.

gap-filling task, and the sentence-writing task, respectively. The same patterns

were found both in the whole group and the intermediate group. As a whole, the descriptive statistics results seem to support the prediction of the involvement load hypothesis across the proficiency levels. In addition, the test scores of the advanced group were always higher than those of the intermediate group in all tasks both in the immediate and the delayed tests, suggesting that the proficiency is an influential factor

Table 4.2.

Effects of Task and Proficiency on the Immediate and the Delayed Tests

Immediate Test						
	<i>SS</i>	<i>Df</i>	<i>MS</i>	<i>F</i>	<i>P</i>	<i>η</i> ²
Task	788.26	2	394.13	49.43	.000*	0.47
Proficiency	318.05	1	318.05	39.89	.000*	0.26
Task x Proficiency	4.42	2	2.21	0.28	.758	0.01
Error	892.95	112	7.97			
Delayed Test						
	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P</i>	<i>η</i> ²
Task	257.16	2	128.58	30.52	.000*	.355
Proficiency	210.24	1	210.25	49.90	.000*	.310
Task x Proficiency	9.24	2	4.62	1.10	.338	.019
Error	467.66	111	4.62			

**p*<.05

In order to check whether the differences are statistically significant, a set of two-way analysis of variance (ANOVA) were conducted. As seen in Table 4.2, both the task type and the proficiency level were found to have the significant main effects on the initial learning and its retention. The immediate test results revealed the main effect of the task, $F(2, 115) = 49.43, p = 0.000$, and the main effect of the proficiency, $F(2, 115) = 39.89, p = 0.000$. The strength of associations was calculated using η^2 . The value was 0.47 for the task type and 0.26 for the proficiency. In the delayed test score analysis, both the task type and the proficiency level had the main effects: $F(2, 114) = 30.52, p = 0.000$, for the task type, and $F(2, 114) = 49.90, p = 0.000$ for the proficiency level. The calculated η^2 values in the delayed post-test were 0.355 for the task type and $\eta^2 = 0.310$ for the proficiency level. In addition, there was no significant interaction effect between the task and the proficiency either in the immediate test ($F = 0.28, p = 0.758$) or the delayed test ($F = 1.10, p = 0.338$).

Post hoc Scheffe's analysis was conducted to discover any significant differences among the task types. Since the proficiency variable had only two levels, the advanced and the intermediate, two-way ANOVA did not provide *post hoc* analysis according to the proficiency. Therefore, an additional set of one-way ANOVAs (one for the immediate test and the other for the delayed test) were conducted for the advanced level group and the intermediate level group, respectively.

Table 4.3

Post Hoc Analysis on the Immediate and Delayed Tests

Proficiency	Task	Task	Immediate Test		Delayed Test	
			Mean	P	Mean	P
			Difference		Difference	
Whole (A+I)	Task1	Task2	-3.09	.000*	-1.33	.019*
	Task1	Task3	-6.45	.000*	-3.67	.000*
	Task2	Task3	-3.37	.000*	-2.34	.000*
Advanced	Task1	Task2	-2.90	.011*	-1.22	.241
	Task1	Task3	-6.58	.000*	-4.10	.000*
	Task2	Task3	-3.68	.011*	-2.88	.001*
Intermediate	Task1	Task2	-3.33	.002*	-1.55	.045*
	Task1	Task3	-6.07	.000*	-3.10	.000*
	Task2	Task3	-2.73	.014*	-1.55	.048*

Note. Task 1 is Answering Questions, Task 2 is Gap-Filling, and Task 3 is Writing Sentences.

* $p < .05$

As shown in Table 4.3, the immediate test results of all three task types were significantly different from each other (all $ps < 0.05$) in all group types (the whole group, the advanced group, and the intermediate group). The mean scores of the

sentence-writing task were significantly larger than those of the gap-filling task, which in turn were significantly better than those of the answering questions task. In the delayed test, the whole group and the intermediate group showed the same results. In the delayed test results of the advanced group, the mean score of sentence writing task was significantly greater than that of the gap-filling task ($p = 0.01$), but the results of the gap-filling task were not significantly better than those of the answering questions task ($p = 2.41$).

4.1.2. The Effect of the Involvement of Load of Tasks on the Active and Passive Word Learning of the Learners at the Different Proficiencies

In this section, the passive word knowledge test score and the active word knowledge test score are treated as two dependent variables and analyzed to investigate whether the involvement load of tasks affects passive word learning and active word learning differently.

Table 4.4 shows the mean scores and the standard deviations of all three groupings (the combination of two groups, the advanced, and the intermediate) in both the immediate and the delayed tests. When it comes to the passive word learning test, the results seem to support the prediction of the involvement load hypothesis both in the long-term and the short-term learning regardless of the

proficiency levels. In all cases, the mean scores were the highest in the sentence writing task, the second highest in the gap-filling task, and the lowest in the answering questions task. The active word learning test results revealed the same patterns for the whole group and the intermediate group in both the immediate and the delayed tests; the mean scores were on the rise in order of task 1, task 2, and task 3. The immediate active word learning test scores of the advanced group were in line with those of the whole group and the intermediate group; the mean scores were 2.22 for task 1, 3.92 for task 2, and 5.61 for task 3. The delayed active word learning scores of the advanced group, however, showed somewhat deviant results; the average score was the highest in the sentence writing task ($M = 2.79$), but the mean score of the gap-filling task with the involvement load index 2 ($M = 1.21$) was smaller than that of the answering questions task with the index 1 ($M = 1.29$).

Table 4.4

Descriptive Statistics of Passive and Active Word Knowledge

Whole Group (High + Intermediate)												
	Immediate Test						Delayed Test					
	Passive			Active			Passive			Active		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Task 1	40	2.24	1.59	40	1.43	1.26	39	1.62	1.36	39	0.88	0.87
Task 2	39	3.51	2.05	39	3.24	1.88	39	2.74	1.97	39	1.08	1.16
Task 3	39	5.64	1.87	39	4.48	1.96	39	4.18	1.57	39	2.00	1.26

Advanced												
Immediate Test						Delayed Test						
Passive			Active			Passive			Active			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Task1	19	3.16	1.45	19	2.22	1.27	19	2.45	1.26	19	1.29	1.01
Task2	18	4.36	1.94	18	3.92	1.45	18	3.75	1.77	18	1.21	1.23
Task3	20	6.35	1.79	20	5.61	1.70	20	5.10	1.39	20	2.79	1.09

Intermediate												
Immediate Test						Delayed Test						
Passive			Active			Passive			Active			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Task1	21	1.40	1.23	21	0.71	0.72	20	0.83	0.94	20	0.50	0.46
Task2	21	2.79	1.89	21	2.67	2.04	21	1.88	1.74	21	0.98	1.12
Task3	19	4.89	1.68	19	3.29	1.46	19	3.21	1.12	19	1.17	0.81

Note. Task 1 is answering questions, Task 2 is gap-filling, and Task 3 is writing original sentences. The maximum score for each test is 10.

The statistical significance of the test result differences was checked through a set of two-way multivariate analysis of variance (MANOVA). As presented in Table 4.5, the task and the proficiency were found to have statistically significant impacts on both the immediate test and the delayed test; *F* values were 22.73 for the task type and 20.55 for the proficiency level in initial learning, and 15.41 and 25.91 in long-term learning. All *ps* were 0.000. As to interaction, only

in the delayed test was there interaction effect ($F=2.87, p=.024$); there was no interaction effect in the immediate test ($F=1.24, p=.295$).

Table 4.5

Effects of Task and Proficiency on the Passive and Active Word Learning

Immediate Test						
	Value	<i>F</i>	<i>df</i>	<i>Error df</i>	<i>P</i>	η^2
Task	0.50	22.73	4	222	0.000*	0.291
Proficiency	0.73	20.55	2	111	0.000*	0.270
Task x Proficiency	0.96	1.24	4	222	0.295	0.022
Delayed Test						
	Value	<i>F</i>	<i>df</i>	<i>Error df</i>	<i>P</i>	η^2
Task	0.61	15.41	4	222	0.000*	0.219
Proficiency	0.67	25.91	2	110	0.000*	0.320
Task x Proficiency	0.91	2.87	4	220	0.024*	0.050

*** $p < .05$**

Post hoc Scheffe's analysis was conducted to detect the significant differences among the task types. Table 4.6 displays *post hoc* analysis results of the immediate passive and active test scores. When it comes to the whole group, both the passive test scores and active test scores demonstrated significant

differences across all three task types; the mean score of task 3(sentence writing) was significantly larger than that of task 2 (gap-filling), which was significantly larger than that of task 1 (answering questions). Regarding the advanced group, the active test scores showed the same results to those of the whole group. Yet, in the passive score test of the advanced group, there was no statistically significant difference between the answering questions task and the gap filling task ($p = 0.119$), although the sentence writing task produced significantly better results than the answering question task or the gap filling task ($p = .000$ and $p = .004$ respectively). The immediate passive word learning test scores of the intermediate group revealed the three tasks significantly differed from each other. Regarding the immediate active word learning test results, however, the sentence writing task did not yield significantly better outcomes than the gap-filling task ($p = 0.430$), though both the sentence writing task and the gap-filling task were significantly more effective than the answering questions tasks (both $ps = .000$).

Table 4.6
Post Hoc Test on the Immediate Test

Proficiency	Task	Task	Passive			Active		
			Mean-Difference	Std. Error	P	Mean-Difference	Std. Error	P
Whole Group (H+I)	Task1	Task2	-1.28	.38	.004*	-1.81	.34	.000*
	Task1	Task3	-3.40	.38	.000*	-3.05	.34	.000*
	Task2	Task3	-2.13	.38	.000*	-1.24	.34	.002*
Advanced	Task1	Task2	-1.20	.57	.119	-1.69	.49	.005*
	Task1	Task3	-3.19	.56	.000*	-3.39	.48	.000*
	Task2	Task3	-1.99	.56	.004*	-1.70	.48	.004*
Intermediate	Task1	Task2	-1.38	.50	.028*	-1.95	.47	.000*
	Task1	Task3	-3.49	.51	.000*	-2.56	.48	.000*
	Task2	Task3	-2.10	.514	.001*	-.62	.48	.430

Note. Task 1 is answering questions, Task 2 is gap-filling, and Task 3 is writing sentences.

* $p < .05$

Table 4.7

Post Hoc Test on the Delayed Test

Proficiency	Task	Task	Passive			Active		
			Mean-Difference	Std. Error	P	Mean-Difference	Std. Error	P
Whole Group (H+I)	Task1	Task2	-1.13	.32	.003*	-.20	.22	.673
	Task1	Task3	-2.56	.32	.000*	-1.12	.22	.000*
Advanced	Task2	Task3	-1.44	.317	.000*	-.92	.22	.000*
	Task1	Task2	-1.30	.49	.035*	.08	.37	.976
	Task1	Task3	-2.65	.47	.000*	-1.50	.36	.000*
Intermediate	Task2	Task3	-1.35	.48	.025*	-1.58	.36	.000*
	Task1	Task2	-1.05	.41	.046*	-.47	.26	.207
	Task1	Task3	-2.39	.42	.000*	-.67	.27	.055
	Task2	Task3	-1.33	.42	.009*	-.20	.27	.769

*p<.05

Note. Task 1 is Answering Questions, Task 2 is Gap-Filling, and Task 3 is Writing Sentences.

Table 4.7 summarizes *post hoc* analysis results of the delayed test. The delayed passive test results of all three groupings (the whole group, the high, and

the intermediate) correspond to the involvement load hypothesis; the sentence writing task was always significantly better than the gap-filling task, which was also always better than the answering questions task (all $ps < .05$). When it comes to the delayed active test scores, the high and the whole group results displayed that the sentence writing task was significantly more effective than the other two tasks (all $ps = .000$), but there were no significant differences between task 2 (gap-filling) and task 1 (answering questions); $p = 0.673$ for the whole group, $p = .976$ for the high group. Concerning the delayed active test score of the intermediate group, there was no significant differences among three tasks (all $ps > 0.055$).

4.2. Discussion

In Section 4.2.1, the impact of the task involvement load on the overall vocabulary learning of the learners at the different proficiency levels is discussed. The discussion on the influences of the task involvement load on the passive word learning and the active word learning is presented in Section 4.2.2.

4.2.1 The Effect of the Task Involvement Load on the Immediate Vocabulary Learning and its Retention of the Learners at the Different Proficiency Levels

The statistical analysis results showed that both the task type and the proficiency level had significant impacts on the overall vocabulary learning both in the short-term and the long-term learning; all *ps* were 0.000. The *post hoc* Scheffe analysis of the immediate post-test provided full support for the involvement load hypothesis; in all three groupings, task 3 was always significantly more effective than task 2, which in turn yielded significantly better results than task 1. In the delayed post-test, however, it provided near complete support for the hypothesis; there was no significant difference between task 1 and task 2 in the advanced level group ($p= 0.241$). But in the whole group and the intermediate group, the involvement hypothesis successfully predicted the learning outcomes of three tasks. Therefore, it can be concluded that the involvement load hypothesis had a good predictive power for the overall learning outcomes of the immediate learning and its retention. This result is in line with other studies, which provided complete or near complete support for the hypothesis (Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008).

Regarding the proficiency, as in the previous research (Kim, 2008), the present study provided supporting evidence for Hulstijn and Laufer's (2001) claim; learners at the different proficiencies can benefit from more involving tasks as long as they can complete the tasks. Not only the advanced learners but also the intermediate learners in the present study performed better in the initial learning and its retention when the task required more mental efforts from the learners.

The fact that the results of both proficiency groups corresponded to the hypothesis may be attributed to the considerable efforts in material modification to ensure the task manageability of the less proficient learners. Except for 10 target words, the vocabulary difficulty and the sentence complexity were geared toward the intermediate level learners. In the pilot study, the focal checkpoint was on the intermediate learners' task manageability. Such effort rarely existed in the previous research which concluded that some tasks with higher involvement load were beneficial only for the high-level learners (Kim and Na, 2010; Sung, 2013). For example, Kim and Na (2010), who conducted the experiment on the university students in Korea at three different levels (the advanced, the intermediate, and the beginner) adopted the exact same text and tasks from Kim's (2008) study, whose participants were the quite advanced learners in the ESL context. The results that only the advanced group was significantly different from the rest groups could have been due to the intermediate and the beginner learners' failure to complete the tasks. Since less proficient learners' capacity to do the task was ensured beforehand in this study, it can be concluded with higher certainty that the learners at the various levels can benefit from the more involving tasks once the threshold for task completion is reached. The results of this study may serve as the ground for the instructional decisions to engage less proficient learners in tasks with higher involvement load with careful task adjustment.

This study is in line with Kim's (2008) study in that the less proficient

learners in both studies could benefit from more involving tasks. Those two studies, however, are distinguished from each other in terms of the significance of proficiency factor. Kim (2008) concluded that the proficiency was insignificant; the mean scores of the two proficiency level groups were similar. In contrast, the proficiency factor was found to be significant in the current study; the high level learners always showed significantly better results than the intermediate group although both groups benefited more as the involvement load of the task increased.

This difference in the significance of proficiency can be explained by the careful classification of the proficiency groups in this study. The insignificance of proficiency in Kim's (2008) study might have come from the problematic proficiency labeling as admitted by Kim herself and criticized by Kim and Na (2010). In Kim's research, the advanced group consisted of international graduate students in the United States and the intermediate group was composed of international undergraduate students enrolled in the intensive language program. The latter group had TOEFL scores around 520, so it might have been problematic to label them as the intermediate learners (Kim, 2008; Kim & Na; 2010). The present study tried to make sure two proficiency groups were clearly different from each other, by placing a 10 % gap between two levels. The superiority of the high learner's performance proved the clear discrepancy between two proficiency groups.

Unlike the whole group and the intermediate group, which showed full

support for the involvement load hypothesis both in the immediate and the delayed post-tests, the high level displayed complete support only for the immediate test. In the delayed test, task 2 (gap-fill) was not significantly more effective than task 1 (answering the comprehension questions). This pattern of the advanced learners is similar to the results of the quite advanced learners in Kim's (2008) research and the Dutch group in Hulstijn and Laufer's (2001) study. In their studies, only one of immediate or delayed post-tests provided complete support for the hypothesis and the other yielded partial support; there were no significant differences between the answering questions task and the gap-filling task in those studies.

The insignificant difference between task 1 and task 2 of the advanced learners of the present study in the delayed post-test may be interpreted in the following two ways; 1) unequal interval among prominences of evaluation component, 2) exceptionally high retention rate in task 1 of the advanced group due to their familiarity with this type of task, which might have paved the way to their active strategy uses.

Firstly, despite Hulstijn and Laufer's (2001) suggestion that every component and its prominence should be assigned the same value, strong evaluation (index load 2) might have produced much more involvement load than moderate evaluation (index load 1). In other words, the interval between strong evaluation and moderate evaluation might not be equal as the interval between the

moderate evaluation and no evaluation. This has been pointed out in the previous research, and the suggestions for more studies to test and revise the hypothesis were made on this ground (Kim, 2008; Park, 2011).

A second possible explanation is the high retention rate of the advanced group in task 1, answering the comprehension question task. The retention rate, the percentage of the words remembered after a week among the words learned right after the task completion, was calculated by dividing the delayed test scores by the immediate test scores. The outcomes were 70% for task 1 (71.1/102.25), 60% for task 2 (83.75/140.5), and 67% for task 3 (67.75/100.5). Even though the involvement load of task 1 was the smallest, the retention rate of the task 1 was the highest. This high retention rate might account for why the significant difference between task 1 and task 2 ($p = 0.000$) in the immediate test disappeared in the delayed test ($p = 0.241$).

This high retention rate of task 1 of the advanced learners might have resulted from the learners' familiarity with this type of task and high level learners' more outstanding learning skills and strategies than the intermediate learners'. In Korean English education, especially in the third grade of high school, the majority of the class time is spent on reading the text and answering the comprehension questions, just as in task 1. While engaging in these tasks continuously in their regular class and self-study, the more proficient learners might have been able to develop better skills in learning vocabulary, contributing to better retention in task

1 than in the unfamiliar task 2. Moreover, task 1 was designed in the same format as the practice CSAT to lessen the cognitive burden on the learners. The familiarity of the task 1 type to the learners is backed up by the results of the questionnaire used as the distracter in this study; the 88.1 % of the learners reported that they allocate most of their English study time on reading for answering questions and picking up new vocabulary, which is directly related to task 1.

The high retention rate of only the advanced learners, despite the familiarity of task 1 to both the advanced and intermediate groups, might be accounted for by the superior learning skills and strategies of the advanced learners. The fact that the more proficient learners develop better skills and strategies not only in general learning but also in vocabulary learning was discovered in many studies. For example, Zahar, Cobb, and Spada (2001) found out that the advanced learners picked up new words in fewer encounters than the less advanced learners, showing the superior learning ability of the higher proficiency learner group. Similar results were constantly reported in the vocabulary learning studies (Kojic-Sabo & Lightbown, 1999; Fan, 2003; Wei, 2007; Yang, 2007).

Since the test scores of the intermediate learners provided full support for the hypothesis, the insignificant difference between task 1 and task 2 of the advanced learners might be better accounted for by the second explanation, which is focused on the characteristics of the advanced learners, than the first interpretation, the suggestion for revision of the hypothesis itself.

To sum up, regarding the first research question, the influence of task-induced involvement load on the overall vocabulary learning of the learners at the different proficiency levels, the involvement load hypothesis was fairly successful in predicting the test results, providing near full support for the hypothesis. The research results also corroborated Hulstijn and Laufer's (2001) claim that less proficient learners can also benefit from the tasks with higher involvement load if they can manage the tasks; both the advanced level learners and the intermediate level groups performed significantly better with the more engaging tasks even though the scores of the advanced group were higher than those of the intermediate group in all comparisons.

4.2.2. The Effect of the Task Involvement Load on the Active and Passive Word Learning of the Learners at the Different Proficiencies

Vocabulary learning has many complex aspects and requires the different approaches depending on the focus (Schmitt, 2008). Except for very few studies (Keating, 2008), however, most empirical studies on the involvement load hypothesis have investigated only one aspect of it, word meaning. Therefore, the need to test the predictive power of the hypothesis on the different aspects of word learning was raised (Kim, 2008). To check whether the involvement load of tasks influences word meaning learning and word form learning differently, a set of

MANOVAs were conducted with the passive word learning test scores (meaning) and the active word learning scores (form) set as two dependent variables. The MANOVA results revealed that both the task type ($p= 0.000$, $\eta^2 = 0.291$ for the immediate test and $p= 0.000$, $\eta^2= 0.219$ for the delayed test) and the proficiency level ($p= 0.000$, $\eta^2 = 0.270$ for the immediate test and $p= 0.000$, $\eta^2= 0.320$ for the delayed test) were the significant factors.

First, when it comes to the passive word learning, the results of the present study verified the involvement load hypothesis. In the whole group and the intermediate group, the sentence writing task (involvement load index 3) was significantly superior to the gap-filling task (index 2), and the latter was significantly better than the comprehension question task (index 1) both in the short-term and the long-term learning. The high level group also mostly satisfied the prediction of the hypothesis with only one exception. Between the comprehension question task (1) and the gap-filling task (2) in the immediate passive test scores, the mean score was higher in the latter task, but the difference was not statistically significant enough ($p=0.119$).

As already discussed in Section 4.2.1, this deviance might be attributed to the high level learners' impressive performance in task 1, which resulted from their daily practice in the task 1 type and their superior learning skills and strategy uses. Another possible explanation may be the unequal distance in prominence of evaluation component. Despite the claim that each index should be assigned the

same value, the strength difference between strong evaluation (2) and moderate evaluation (1) might be much bigger than the gap between moderate evaluation (1) and no evaluation (0) as also suggested in other studies (Laufer & Hustijn, 2001; Kim, 2008; Park; 2011). In either case, from the fact that the main focus of the involvement load is on the long-term retention, it can be concluded that this research provided the strong support for the involvement load hypothesis, at least in terms of the passive word learning; the delayed test results perfectly corresponded to the prediction of the hypothesis.

In contrast, with regard to the active word learning test scores, the involvement load hypothesis quite successfully predicted the learning outcomes in the initial learning, but failed to do so in long-term learning. The immediate active word test scores corresponded to the prediction of the involvement load hypothesis across all three proficiency groups (all $p < 0.05$) with only one exception; task 2 (gap-filling) and task 3 (sentence writing) of the intermediate group did not show the significant difference ($p = 0.43$). In the delayed active word learning test, however, only task 3 was significantly more effective than task 1 or task 2 (all $p = 0.00$) in the whole group and the advanced group; there existed no further significant differences among the tasks in those two groups. The delayed active test results of the intermediate group revealed that three tasks were not significantly different from each other at all (all $p > 0.055$). In short, in terms of the delayed active word learning test, there was only the superiority of task 3

(sentence writing), and this superiority was effective only in the high level group. The interaction effect found in the delayed test scores ($p=0.024$) seems to have resulted from this rather complex pattern of the delayed active test scores according to the proficiency level.

To argue for the predictive power of the involvement load hypothesis in the immediate active word learning, it should be explained why there was no significant difference between task 2 (index 2) and task 3 (index 3) only in the intermediate learners. The insignificant differences in the word form test scores of the intermediate learners between task 2 and task 3 may be attributed to the intermediate learners' limited mental capacity in processing L2; the less proficient learners might not be able to attend both to the meaning and the form simultaneously, if the task demand is high for their competence (Vanpatten, 1990; Barcroft, 2002). Vanpatten (1990) tentatively suggested that for the learners to be able to pay attention to the forms, the task and the material must be easy and comprehensible. Similarly, Barcroft (2002) proposed TOPRA model (type of processing-resource allocation), which predicts that the semantically elaborative task facilitates only the word meaning learning but decreases the possibilities for the form and structure learning, if learners have limited mental capacity.

When the less proficient learners are not able to attend to both form and meaning, whatever mental capability left for word learning is more likely to be directed to meaning (i.e., passive word learning) than to form (i.e., active word

learning) since the former is easier than the latter. It is often pointed out that learning a word productively is more difficult than learning a word receptively (Nation, 1990). Laufer and Goldstein (2004) also placed the recalling word form higher in the hierarchy of word learning difficulty than the recalling word meaning. On top of that, the learners in this study were instructed not to pay too much attention to the grammar or accuracy but to make their message across in task 3. This direction might have magnified the learners' tendency to choose to pay attention to word meaning over word form when attending to both is beyond their mental capabilities.

For the advanced learners, all three tasks yielded the immediate active word test scores as predicted by the hypothesis, since all three tasks were manageable enough for them to turn their attention to form as well as meaning. For the intermediate learners, however, task 1 and task 2 might have been within their threshold level, but task 3 might have been too difficult for them to focus both on form and meaning.

Korean learners, especially less proficient learners, usually feel more frustration and difficulty when facing the writing task such as in the task 3 type. According to Kwon (2004), the Korean students found the writing task the most challenging among four skills. The writing task scores of the Korean learners were lowest among the three groups, China, Korea, and Japan. The challenging nature of task 3 was also revealed in the questionnaire results as well. Among 258

participants, only two learners reported they practice writing and receive writing instruction during the regular class hour. The self-estimated proficiency was also the lowest in the writing; on the scale from 1 (very good) to 5 (very bad), the average score was 1.75 in reading, 1.96 in listening, 2.99 in speaking, and 3.37 in writing for the advanced group, and it was 2.43, 2.36, 3.20, and 3.97 for the intermediate group. The questionnaire results about the learners' perception of the difficulty also revealed that the learners, especially the intermediate learners, felt the most challenged in task 3. On the Likert scale from 1 (very difficult) to 5 (very easy), the average for the advanced level group was 4.12, 3.51, and 2.81 for task 1, 2, 3, respectively. For the intermediate group, it was 3.56, 2.68, and 1.15.

In short, the immediate active word learning outcomes in this study can be explained by the involvement load hypothesis. The only insignificant difference between task 2 and task 3 of the intermediate group was probably due to the more challenging nature of task 3. Since task 3 was difficult for their ability, the intermediate learners might have been able to focus only on meaning, but not on form, while task 3 may have been manageable enough for the advanced learners to pay attention not only to meaning but also to form.

The predictive power of the involvement load hypothesis on the active word learning, however, became invalid in the delayed active vocabulary test scores, showing more complex pattern depending on the proficiency level. The superiority of task 3 was found only in the advanced group, and there were no

significant differences among three tasks in the intermediate group. This superiority of task 3 can be interpreted in two ways; 1) the involvement load hypothesis cannot explain the retention of word form learning, 2) the retention of word form learning might be more related to the actual processing time of target vocabulary.

The retention of the active vocabulary learning might be explained better by other theories than the involvement load hypothesis. One possible candidate is ‘transfer appropriate processing theory’ (Morris, Bransford, & Franks, 1977), which focuses on the interaction between the nature of the tasks and the test types. The involvement load hypothesis is based on ‘processing depth theory’, which claims that the semantically deep processing is more conducive to building lasting memory traces than shallow form-related elaboration such as processing grapheme (Criak & Lockhart, 1972). Transfer appropriate processing theory, however, recognizes the value of different levels of processing, and argues that the value of the task can only be determined by the purpose of the task, not merely by how much semantically involving the task is. In other words, the most effective task for the recall of meaning might not be the most optimal activity for the recall of other aspects of words such as spellings (Tulving & Thomson, 1973; Morris et al., 1977; Barcroft, 2002). For example, in Morris, Bransford, and Frank’s (1977) experiment, the participant performed better in the word meaning recognition test when they completed semantically elaborative task than when they engaged in the

rhyme task. When the given measurement was a rhyme recognition test, however, the rhyme elaboration task was found to be more effective than the semantically involving task.

Whereas task 1, reading and answering the comprehension questions, and task 2, reading and choosing the most appropriate words from the context, are basically comprehension-based activities, only task 3, writing original sentences, is production-based activities; there is much higher possibility for the learners to pay attention to word form in task 3 than in task 1 or 2. In other words, while the task 1 and 2 were mostly semantically elaborative tasks, task 3 was structurally as well as semantically involving task. Since all three tasks were semantic elaboration tasks, the passive word learning test produced the results which fit the involvement load hypothesis frame, which is based on the semantic-associated theory. When the test type is more form-oriented, however, only task 3, the only one with the strong form components, seems to have the distinctive advantage. Therefore, only active word learning attained through task 3, where the task type and the measurement type matched, remained in the delayed test.

This superiority of task 3 in the delayed active word test manifested only in the advanced learners, since only they have enough L2 processing capability to focus both on form and meaning in the challenging task 3. In contrast, the intermediate learners, who had limited mental resource left for the target vocabulary, could not focus on form when the task was too challenging; for the

intermediate group, task 3 was not significantly more effective than the other tasks in both the immediate and delayed active word learning test.

Another possible explanation of the superiority of task 3 in the delayed active word learning is that the retention of the formal property of words is more dependent on how long and how many times the items were processed. In fact, the number of the words in each task, input, was the highest in task 1 (476 words), the second highest in task 2 (528 words), and the lowest in task 3 (192 words). The order of the task completion time was reversed; the learners spent the longest time in task 3 (756 seconds on average), then the second longest in task 2 (512.8), and the least in task 1 (280.6). When the time spent on per word was calculated by dividing the average task completion time by the number of words to process, it was 0.59, 0.94, and 4.21 for task 1, 2, and 3, respectively. The learners in the present study might have processed and repeated the target word much more times and much longer in task 3 than in the task 1 or 2, and this might have resulted into the superior retention of word form in task 3.

The two possible interpretation of the superiority of task 3 in the retention of the active word learning might have worked together, producing the synergy effects. The learners could benefit in the word form learning only when the task itself has the form elaboration component as in task 3, and this effect might have been magnified by that they were able to process the target words more times and

for longer period when engaged in task 3.

To conclude, the involvement load of the tasks influenced the passive word learning and the active word learning differently. The passive word recall, the semantic-based test, provided support for the involvement load hypothesis, which is also based on the semantic theory, the processing depth theory. But when the focus of the measurement shifts to the other aspects, i.e., the word form, the results were not consistent with the involvement load hypothesis. In this case, whether the task type matches the test type might have been more critical. Therefore, the results of the present study indicate the involvement load hypothesis might be confined to the semantic aspects of the word learning and may not encompass all the aspects of vocabulary learning. Hultstin and Laufer (2001) claimed that higher involvement will lead to longer retention regardless of task types, but the results of the present study suggest that the type of the task might matter. Certain task types might be more beneficial when learning is measured by a certain method, since different types of tasks can lead learners to pay attention to the different aspects of vocabulary. Therefore, caution might be necessary in judging and predicting the effectiveness of the tasks in terms of the involvement load of the tasks. The instructors and learners might need to think of the purpose of learning first when they judge or compare the effectiveness of the vocabulary tasks.

Chapter 5.

CONCLUSION

This chapter is composed of three sections. Section 5.1 summarizes the major findings of the present study. In Section 5.2, the implications are presented on English vocabulary education and on the involvement load hypothesis. Finally, Section 5.3 reports the limitations of the present study and makes suggestions for the further research.

5.1 Major Findings

This study investigated the effects of the task-induced involvement load on English vocabulary learning of Korean high school learners at the different proficiency levels. The first research question looked into the effect of the involvement load of the tasks on the overall vocabulary learning and its retention. In the second research question, the effects of the task-induced involvement load on the passive vocabulary learning were compared with those on the active vocabulary learning.

From the analysis of the sum of the passive vocabulary test scores and the

active vocabulary test scores, the present study provided near complete support for the involvement load hypothesis; the higher the task-induced involvement load was, the better the learners performed both in the immediate test and the one-week delayed test. This strong support for the involvement load hypothesis was demonstrated not only in the advanced level group, but also in the intermediate group. This result is consistent with the claim by Laufer and Hulstijn (2001) that learners at various proficiency levels can benefit from tasks with higher involvement loads if the task is manageable to the learners.

When passive word learning and active word learning were separately analyzed, they showed the different patterns. Regarding the passive word learning, the results corresponded to the prediction of the involvement load hypothesis. But the active word learning test results, specifically the long-term learning, could not be explained by the involvement load hypothesis. Only task 3 was found significantly more effective among three tasks, and this superiority of task 3 was found only in the advanced learners.

The effectiveness of task 3 (writing original sentences) in the retention of active word learning might be accounted for better by other theories, such as the transfer appropriate processing theory (Morris et al., 1977). The involvement load hypothesis and processing depth theory (Craik & Lockhart, 1972), which the hypothesis was based on, assume that information is retained better when it is processed in a semantically deep level. In contrast, the transfer appropriate

processing theory recognizes the interaction between the task type and the measurement type; the semantically involving task is not always the optimal task depending on the test type. In the current study, only task 3 had the potentially strong form factor. Therefore, when the test focused on the word form as in the active word learning test, only task 3 was found significantly effective. The alternative explanation is that the word form learning is more proportionate to how many times and how long the target forms were processed; the calculated time spent on the target vocabulary was much longer in task 3 than in task 1 and 2. This superiority of task 3 may be valid only when the learners can complete the task easily; the superiority of task 3 was found only in the advanced learners.

Overall, the present study provides quite strong evidence for the involvement load hypothesis; the immediate and delayed passive word learning scores and the immediate active word learning scores mostly corresponded to the prediction of the hypothesis. Moreover, the prediction was valid not only in the advanced group but also in the intermediate group, confirming Laufer and Hustijn's (2001) claim that the involvement load hypothesis can explain word learning of the learners at the various proficiency levels if the task is manageable. Regarding the retention of the active word learning, however, only task 3 was effective for the advanced learners, but not for the intermediate learners, indicating the involvement load hypothesis might not be able to explain all the aspects of vocabulary knowledge, for example, word form learning.

5.2 Implications

Based on the major findings described in session 5.1, this study presents the following pedagogical implications on L2 vocabulary learning.

- 1) Implementing more cognitively demanding vocabulary tasks in the classroom context, in which the learner proficiency gap exists, may be effective for vocabulary learning since the intermediate learners as well as the high level learners can benefit from the tasks with higher involvement load.
- 2) Rather than just assuming the semantically deep task is always the most effective, the effectiveness of the tasks should be judged based on the educational goal of the vocabulary tasks. For example, if the focus is on the word form, the task which involves semantically shallow but structurally intensive processing of new words may be the most effective. The teachers should ask what aspects of the vocabulary learning they want to promote, before designing, modifying, comparing, choosing, or implementing vocabulary tasks.
- 3) It is crucial to make the vocabulary tasks manageable to the target learners. When the tasks are beyond the learners' threshold levels, the tasks cannot be useful, no matter how effective the task was proven to

certain learners. To make the effects of the task with the high involvement loads valid, the practitioners should make proper modifications to the tasks or make efforts to help the learners be prepared for the tasks.

Beside the pedagogical implications, the results of the this research provide implications regarding the involvement load hypothesis.

- 1) Although the research findings generally provide full support for the hypothesis in overall word learning and passive word learning, there were a few cases inconsistent to the prediction of the hypothesis. In the present study, the insignificant differences in the test results always occurred between task 1 (no evaluation) and task 2 (modest evaluation). Similar patterns were found in some previous studies (Hulstijn & Laufer, 2001; Kim, 2008). Those findings might suggest the need for revising the involvement load scale, at least between the ‘no evaluation’ and ‘modest evaluation.’
- 2) The predictive power of the involvement load was only valid in the passive word learning, more related with word meaning, but invalid in the active word learning, more related to word form. The involvement

load hypothesis might be confined to the retention of the word meaning and might not be able to explain every aspect of vocabulary knowledge.

5.3 Limitations and Suggestions

This study has some limitations despite the fact that each experiment step was meticulously designed and the variables were carefully controlled.

Firstly, the participants of the present study were sampled from a considerably privileged school. The school was placed 107th out of 2,282 high schools in Korea based on the number of the students within the top 11% in the 2013 CSAT. In addition, almost one out of 4 learners turned out to have experience of studying English abroad. Although the school is a public high school, such status is not so common in Korea. In addition, the learners were very cooperative, passionate, interested in English, and even competitive to some extent in task completion. Therefore, the generalization of the findings of the present study might warrant caution. The studies in the schools consisting of different or more diverse student background within the EFL context might be needed. In addition, due to the small number of low level learners and their high absent rate, only the high and the intermediate level learners were compared. Future studies can replicate the experiment in a Korean high school, which has a more balanced

proficiency makeup so that the impact of the involvement load of tasks on more diverse proficiency groups can be compared.

A second limitation is that the scope of active word learning focused in the study was quite narrow. In this study, only the word form, specifically the spelling, was tested as a part of active word learning. This decision was made since the learners were exposed to the target word only once in the task stage. It seemed unrealistic to expect the learners to produce the words properly in context after only one instance of exposure, especially in task with low involvement load. Further research can investigate various aspects of productive word knowledge, for example, using the words in the context after multiple exposures to the target words. Studies on passive word learning can be also expanded to include collocations, figurative meaning of the words, and so on. Since the present research suggests that the involvement load might not cover all aspects of word knowledge, the investigation of various aspects of the vocabulary will be especially valuable.

Next, task 3 in this study, writing original sentences, was the only task with the strong form component. But at the same time, it was the task with the highest involvement load index. Therefore, it is rather hard to say for certain that the learning of formal properties of words is not influenced by the involvement load of task. Further studies can look into active vocabulary learning by designing tasks more systematically, for example, by including task in which the vocabulary

item is processed semantically at a shallow level but intensively at a structural level.

Lastly, even though the present result provided strong support for the involvement load hypothesis in the overall word learning and passive word learning, considerable deviant results existed, specifically between task 1 (answering the comprehension questions) and task 2 (gap filling). This might indicate the need to modify the involvement load component system, especially between no evaluation and modest evaluation. Therefore, the future studies might want to use different tasks in testing out the hypothesis. It should be kept in mind, however, that the new tasks should be carefully designed according the involvement load hypothesis and the other external variables should be controlled with care; the implementation of the thoroughly-designed task with careful variable controls usually yield the corroborating evidences. In addition to the evaluation factor, the exploration of other factors and prominence degrees of them, such as need and search, can also be helpful in refining the hypothesis.

REFERENCES

- Anderson, R.C., & Freebody, P. (1981). Vocabulary and Knowledge. In J. T. Gutrie (Ed). *Comprehension and Teaching: Research Review* (pp. 77-117). Newark, DE: International Reading Association.
- Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 85, 139-152.
- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52(2), 323-363
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, 7(1), 15-32.
- Candlin, C. N. (1988). Preface. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching*. New York: Longman
- Cho, K. S., & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, 37(8), 662-667.
- Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225-237). Cambridge: Cambridge University Press.

- Cohen, R. F., & Miller, J. L. (2004). *NorthStar Reading and Writing, Advanced*.
London: Longman
- 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., & Bamford, J. (1998). *Extensive reading in the second language
classroom*. New York: Cambridge University Press.
- Day, R. R., Omura, C., & Hiramatsu, M. (1992). Incidental EFL vocabulary
learning and reading. *Reading in a foreign language*, 7, 541-541
- Day, R. R., & Swan, J. (1998). Incidental learning of foreign language spelling
through targeted reading. *TESL Reporter*, 31(1), 1-9.
- de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary.
Studies in second language acquisition, 24(01), 81-112
- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental
acquisition of word meanings. *Studies in Second Language Acquisition*,
21(02), 285-301.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction,
comprehension, and the acquisition of L2 word meanings. *Language*

learning, 44(3), 449-491

- Eysenck, M. W. (1982) Incidental learning and orienting tasks. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 197-228). New York: Academic Press.
- Fan, M.Y. (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *Modern Language Journal*, 87(2), 222-241.
- Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in second language acquisition*, 21(02), 225-241.
- Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching*, 41(2), 87-106.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207–223
- Hsueh-Chao, M.H., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.
- Huang, S., Willson, V., & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal*, 96(4), 544-557.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings:

- Experiments in incidental vocabulary learning. In P. J. Arnaud & H. Bejoint (Eds.): *Vocabulary and Applied Linguistics* (pp. 113-25). London: Macmillan.
- 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., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327-339.
- 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, 1-16
- Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied linguistics*, 19(3), 357-377.
- Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*,

12(3), 365-386.

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

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language learning*, 58(2), 285-325.

Kitajima, R. (2001). The effect of instructional conditions on students' vocabulary retention. *Foreign Language Annals*, 34(5), 470-482.

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

Koda, K. (1989). The effects of transferred vocabulary knowledge on the development of L2 reading proficiency. *Foreign language annals*, 22(6), 529-540.

Kolers, P. A. (1975). Specificity of operations in sentence recognition. *Cognitive Psychology*, 7(3), 289-306.

Kojic-Sabo, I., & Lightbown, P. M. (1999). Students' approaches to vocabulary learning and their relationship to success. *Modern Language Journal*, 83(2), 176-192.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The modern language journal*, 73(4), 440-464.

- Kwon, Oryang. (2004, March). *Benesse Corporation's comparative study of the English ability of Chinese, Korean and Japanese high school students' interpretations and considerations from a Korean perspective*. Paper presented at the Benesse Corporation Conference, Seoul, Korea.
- Laufer, B. (1989). 'What percentage of text-lexis is essential for comprehension?' In C. Lauren & M. Nordam (Eds.), *Special language: From humans thinking to thinking machines*, (pp. 316-323). Clevedon: Multilingual Matters.
- Laufer, B. (2000). Electronic dictionaries and incidental vocabulary acquisition: Does technology make a difference? In U. Heid, S. Evert, E. Lehmann, & C. Rohrer(Eds.), *EURALEX* (pp. 849–854). Stuttgart: Stuttgart University Press.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5, 223-250.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.
- Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it?. *RELC journal*, 28(1), 89-108.

- Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language learning*, 57(1), 87-118.
- Lee, Y.H. (2006). *The impact of Task-induced Involvement Load on EFL vocabulary Learning*, unpublished master's thesis, Seoul National University, Seoul, Korea.
- Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language learning*, 43(2), 263-279
- McDaniel, M.A. (1984). The role of elaborative and schema processing in story memory. *Memory and Cognition*, 12, 46-51
- Moeser, S. D. (1983). Levels of processing: Qualitative differences or task demand differences? *Memory and Cognition*, 11, 316-323.
- 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.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading research quarterly*, 19(3), 304-330.
- Nagy, W. E., Anderson, R. C., & Herman, P. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237-270
- Nassaji, H. (2002). Schema theory and knowledge-based processes in second

- language reading comprehension: A need for alternative perspectives.
Language Learning, 52(2), 439–482
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Heinle & Heinle.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening?
Canadian Modern Language Review, 63(1), 59-82.
- Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–200). Cambridge: Cambridge University Press.
- Paribakht, T. S., & Wesche, M. (1999). Reading and “incidental” L2 vocabulary acquisition. *Studies in second language acquisition*, 21(02), 195-224.
- Park, J. H. (2011). *The Role of Task-Induced Involvement Load in Vocabulary Acquisition of Korean College Students*. Unpublished master's thesis. Ehwa Womans' University, Seoul, Korea.
- 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
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language teaching research*, 12(3), 329-363.
- Sung, H. M. (2013). *The effects of task-induced involvement load on EFL high school learners' vocabulary learning*. Unpublished doctorate's dissertation. Sookmyung women's University, Seoul, Korea.
- Tekmen, E., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, 39(2), 220-243.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, 80(5), 352.
- VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287-301
- 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.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65.
- Wei, M. (2007). An examination of vocabulary learning of college-level learners of English in China. *Asian EFL Journal*, 9(2), 93-114.

- Xiaohui, X. (2010). An empirical study on the effect of task on L2 incidental vocabulary acquisition through reading. *Asian Social Science*, 6 (7), 126-131.
- Yang, M. N. (2007). Language learning strategies for junior college students in Taiwan: Investigating ethnicity and proficiency. *Asian EFL journal*, 9(2), 35-57.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, 57(4), 541-572.
- Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. N. Huckin (Eds.), *Second language vocabulary acquisition*, 5-19. New York: Cambridge University Press.

Appendix 1. Pretest

아이디 _____

아래의 단어 중에서 자신이 알고 있는 단어가 있다면 표시(√)하고
그 단어의 뜻을 적어주세요.

- | | |
|---------------------------------------|---|
| <input type="checkbox"/> flank : | <input type="checkbox"/> fanatical : |
| <input type="checkbox"/> liling : | <input type="checkbox"/> mothball : |
| <input type="checkbox"/> plead : | <input type="checkbox"/> exasperation : |
| <input type="checkbox"/> repressed : | <input type="checkbox"/> raunchy : |
| <input type="checkbox"/> infuriate : | <input type="checkbox"/> ideography : |
| <input type="checkbox"/> gibberish | <input type="checkbox"/> dissociate : |
| <input type="checkbox"/> blotch : | <input type="checkbox"/> stoical : |
| <input type="checkbox"/> musty : | <input type="checkbox"/> twitch : |
| <input type="checkbox"/> smattering : | <input type="checkbox"/> cluck : |

Appendix 2. Task 1

아이디 _____ // 시작 시간 _____시 _____분

※ 시작 시간과 끝난 시간을 반드시 기입하세요.

[1-4] 다음 글을 읽고, 물음에 답하시오.

It's still there, the Chinese school where my brother and I used to go.

Every day at 5 P.M., my brother and I had to go to Chinese school. My mother was so determined to have us learn Chinese, the language of our heritage. Kicking, screaming, or crying did not stop my mother.

The **musty** classroom smelled like Chinese medicine. I preferred fresh perfume scents of my American teachers in public school. There was a stage, **flanked** by an American flag, and a Chinese flag. I thought the American flag was prettier.

The emphasis at school was mainly learning Chinese. I had to write **ideographs** carefully with a real ink pen. The pen had to be held in an awkward way if you want to avoid **blotches**.

I felt I had better things to learn than Chinese. Actually, the language was a source of embarrassment to me. Often, I tried to separate myself from my grandmother's unpleasantly loud voice whenever we went out together. She was a little woman in her 70s, but she could shout more sharply and loudly than anyone. Her humor was **raunchy** and her Chinese was rhythmless, quick, and loud. It was not beautiful. It was not like the quiet, **lilting** sound of French.

I did not want people to think I am talking **gibberish**. When I speak English, people nodded at me, smiled sweetly. Even the Chinese people would **cluck** and say I would do well in the world outside Chinatown.

My brother was even more passionate than I about speaking English. He was especially hard on my mother, for her **smatterings** of English. "It's not 'What it is,' Mom," he would say angrily. "It's 'what is it!'" Whenever my mom could not pronounce 'r' correctly, my little brother would stop her in mid-sentence: "Say it again, Mom. Say it right." It **infuriated** my mother.

After two years of struggle with a real ink pen and Chinese words, I was finally permitted to stop Chinese school. I thought of myself as multicultural. I preferred American food to Chinese food; I enjoyed Christmas more than Chinese New Year. At last, I was one of you; I wasn't one of them. Sadly, I still am one of them.

- | | |
|----------------------------------|--|
| 1) musty(형용사) : 곰팡내 나는, 케케묵은 | 7) gibberish(명사) : 황설수설 |
| 2) flank(동사) : ~의 측면에 위치하다. | 8) cluck(동사) : (호의 혹은 불만의 표시로) 혀를 쫓쫓 차다. |
| 3) ideograph(명사) : 표의 문자 | 9) smattering(명사) : 겉핥기 지식, 조금 |
| 4) blotch(명사) : 얼룩 | 10) infuriate(동사) : 극도로 화나게 만들다 |
| 5) raunchy (형용사) : 싸구려의, 선정적인 | (특히 언어지식에 대해) |
| 6) lilting(형용사) : 경쾌하게 오르내리는 억양의 | |

1. 위 글의 제목으로 가장 적절한 것은? [3점]

- ① The Linguistic Differences between Chinese and English
- ② How to Overcome Generation Gaps
- ③ The Importance of English as International Language
- ④ The Struggle to Be an All-American Girl
- ⑤ The Benefits of being Multi-cultural

2. 다음 보기 중 연결 관계가 나머지와 다른 하나는? [2점]

- ① perfume – Chinese medicine ② rhythmless – liltng
- ③ ideograph – blotch ④ Christmas – Chinese New Year
- ⑤ I – my grandmother

3. 다음 중 위 글의 내용과 일치하지 않는 것은? [2점]

- ① 어머니는 자식들이 조상의 언어인 중국어를 배우기를 원했다.
- ② 중국 단어를 손으로 쓰는 것은 쉽지 않은 일이었다.
- ③ 할머니는 누구보다도 말을 크고 날카롭게 했다.
- ④ 할머니의 유머는 상스러웠다.
- ⑤ 어머니는 유창하고 정확하게 영어를 구사했다.

4. 글쓴이(I)에 대한 설명과 일치하는 것은? [2점]

- ① 중국어를 원래 배우고 싶어서 기꺼이 중국어 학교에 갔다.
- ② 중국 국기의 단순함과 강인함에 매료되었었다.
- ③ 학창 시절 할머니와 외출한 적은 거의 없었다.
- ④ 남동생이 어머니에게 상처 주는 것을 보고 남동생과 크게 다투었다.
- ⑤ 시작한지 2년 후 중국어 학교를 그만 두었다.

감사합니다♥ 끝난 시간 _____ 시 _____ 분

Appendix 3. Task 2

아이디 _____ // 시작 시간 _____시 _____분

※ 시작 시간과 끝난 시간을 반드시 기입하세요.

[1-3] 다음 글을 읽고, 물음에 답하시오.

I's still there, the Chinese school where my brother and I used to go.

Every day at 5 P.M., my brother and I had to go to Chinese school. My mother was so determined to have us learn Chinese, the language of our heritage. Kicking, screaming, or crying did not stop my mother.

The 1) _____ classroom smelled like Chinese medicine. I preferred fresh perfume scents of my American teachers in regular school. There was a stage, 2) _____ by an American flag, and a Chinese flag. I thought the American flag was prettier.

The emphasis at school was mainly learning Chinese. I had to write 3) _____ carefully with a real ink pen. The pen had to be held in an awkward way if you want to avoid 4) _____ .

I felt I had better things to learn than Chinese. Actually, the language was a source of embarrassment to me. Often, I tried to separate myself from my grandmother's unpleasantly loud voice whenever we went out together. She was a little woman in her 70s, but she could shout more sharply and loudly than anyone. Her humor was

5) _____ and her Chinese was rhythmless, quick, and loud. It was not beautiful. It was not like the quiet, 6) _____ sound of French.

I did not want people to think I am talking 7) _____. When I speak English, people nodded at me, smiled sweetly. Even the Chinese people would 8) _____ and say I would do well in the world outside Chinatown.

My brother was even more passionate than I about speaking English. He was especially hard on my mother, for her 9) _____ of English. "It's not 'What it is,' Mom," he would say angrily. "It's 'what is it!'" Whenever my mom could not pronounce 'r' correctly, my little brother would stop her in mid-sentence: "Say it again, Mom. Say it right." It 10) _____ my mother.

After two years of struggle with a real ink pen and Chinese words, I was finally permitted to stop Chinese school. I thought of myself as multicultural. I preferred American food to Chinese food; I enjoyed Christmas more than Chinese New Year. At last, I was one of you; I wasn't one of them. Sadly, I still am one of them.

1. 아래의 단어 목록 중 적절한 단어를 선택하여 빈칸을 채워 본문을 완성하시오.

[빈칸 당 1점]

* 선택된 단어들을 단 한 번씩만 사용하시오.

* 하나의 빈 칸에는 한 단어만 들어갑니다.

가) abate (동사) 줄다, 감소시키다	아) infuriate (동사) 극도로 화나게 만들다
나) apprehensive (형용사) 염려하는, 두려워하는	자) ideograph (명사) 표의 문자
다) blotch (명사) 얼룩	차) lilting (형용사) 경쾌하게 오르내리는 억양의
라) caveat (명사) 경고	카) musty (형용사) 곰팡내 나는, 케케묵은
마) cluck (동사) (호의 혹은 불만의 표시로) 혀를 짹짹 차다	타) proliferation (명사) 확산
바) flank (동사) ~의 측면에 위치하다	파) raunchy (형용사) 싸구려의, 선정적인
사) gibberish (명사) 횡설수설	하) smattering (명사) 걸핍기식 지식 조금(특히 언어 지식에 대해)

2. 위 글의 제목으로 가장 적절한 것은? [3점]

- ① The Linguistic Differences between Chinese and English
- ② How to Overcome Generation Gaps
- ③ The Importance of English as International Language
- ④ The Struggle to Be an All-American Girl
- ⑤ The Benefits of Being Multi-cultural

3. 글쓴이(I)에 대한 설명과 일치하는 것은? [2점]

- ① 중국어를 원래 배우고 싶어서 기꺼이 중국어 학교에 갔다.
- ② 중국 국기의 단순함과 강인함에 매료되었었다.
- ③ 학창 시절 할머니와 외출한 적은 거의 없었다.
- ④ 남동생이 어머니에게 상처 주는 것을 보고 남동생과 크게 다투었다.
- ⑤ 시작한지 2년 후 중국어 학교를 그만 두었다.

감사합니다♥ 끝난 시간 _____ 시 _____ 분

Appendix 4. Task 3

※ 아래의 10개의 단어들을 이용하여 한 문장씩 만들어 보려고 합니다.

주어진 예시문을 참고하되, 본인의 문장을 영작해 주세요.

문법이나 철자의 정확성은 중요하지 않습니다. 문장의 뜻을 알 수 있으면 충분합니다. (문법이나 철자에 너무 신경 쓰지 마세요)

각 문장은 가능하면 7개 이상의 단어로 구성해주세요.

1) **blotch**(명사): 얼룩

예) The pen had to be held in the right way to avoid **blotches**.

→

2) **cluck**(동사): (호의 혹은 불만의 표시로) 혀를 찌다

예) Even the Chinese people would **cluck** and say I would do well in the world
outside Chinatown.

→

3) **flank**(동사): ~의 측면에 위치하다

예) There was a stage, **flanked** by an American flag.

→

4) **gibberish**(명사): 황설수설

예) I did not want people to think I am talking **gibberish**.

→

5) **infuriate**(동사): 극도로 화나게 만들다

예) Whenever my little brother would stop her in mid-sentence, it **infuriated** her.

→

6) **ideograph**(명사): 표의 문자

예) I had to write **ideographs** carefully with a real ink pen

→

7) **lilting**(형용사): 경쾌하게 오르내리는 억양의

예) Her Chinese was rhythmless and loud, not like the quiet, **lilting** sound of French.

→

8) **musty**(형용사): 곰팡내 나는, 케케묵은

예) The **musty** classroom smelled like Chinese medicine.

→

9) **raunchy** (형용사): 싸구려의, 선정적인

예) Her humor was not elegant, but **raunchy**.

→

10) **smattering**(명사): 겉핥기식 지식, 조금(특히 언어 지식에 대해)

예) He was especially hard on my mother, for her **smatterings** of English.

→

감사합니다♥ 끝난 시간 _____ 시 _____ 분

Appendix 5. Passive Word Learning Test; Immediate

1차 아이디 -----

※ 다음 영어단어에 해당하는 우리말을 쓰세요.

품사에 맞는 형태로 써주세요. [예-pretty; 예쁜(o) , 예쁘다(X)]

제한 시간은 3분입니다.

- 1) **cluck** →
 - 2) **raunchy** →
 - 3) **blotch** →
 - 4) **infuriate** →
 - 5) **smattering** →
 - 6) **flank** →
 - 7) **gibberish** →
 - 8) **ideograph** →
 - 9) **musty** →
 - 10) **lilting** →
-

감사합니다! 수고했어요 *^^*

이 과제를 통해 배운 단어가 아닌, 원래 알고 있던 단어가 혹시 있다면
표시해 주세요!

<input type="checkbox"/> musty	<input type="checkbox"/> flank	<input type="checkbox"/> ideograph	<input type="checkbox"/> blotch	<input type="checkbox"/> raunchy
<input type="checkbox"/> lilting	<input type="checkbox"/> gibberish	<input type="checkbox"/> cluck	<input type="checkbox"/> smattering	<input type="checkbox"/> infuriate

Appendix 6. Active Word Learning Test; Immediate

1차 아이디어 _____

※ 우리말에 해당하는 영어 단어를 쓰세요.

본문에 나왔던 단어를 써야 합니다.

일부만 써도 좋으니, 최대한 기억나는 대로 써주세요.

제한 시간은 3분입니다.

- 1) ~의 측면에 위치하다 →
- 2) 표의 문자 →
- 3) (호의 혹은 불만의 뜻으로)
혀를 짹짹 차다 →
- 4) 경쾌하게 오르내리는 역양의 →
- 5) 횡설수설 →
- 6) 얼룩 →
- 7) 싸구려의, 선정적인 →
- 8) 곁핥기식 지식,
조금(특히 언어지식에 대해) →
- 9) 곰팡내 나는, 케케묵은 →
- 10) 극도로 화나게 만들다. →

Appendix 7. Passive Word Learning Test; Delayed

아이디 _____

※ 다음 영어 단어에 해당하는 우리말을 쓰세요.

품사에 맞는 형태로 써주세요. [예-pretty; 예쁜(o), 예쁘다(X)]

제한 시간은 3분입니다.

- 1) **lilting** →
- 2) **smattering** →
- 3) **blotch** →
- 4) **Gibberish** →
- 5) **Raunchy** →
- 6) **Flank** →
- 7) **Musty** →
- 8) **Ideograph** →
- 9) **Infuriate** →
- 10) **Cluck** →

감사합니다! 수고했어요 *^^*

저번 실험 후 (7일 전) 한 번이라도 더 공부한 단어가 있다면 표시해주세요.

<input type="checkbox"/> musty	<input type="checkbox"/> flank	<input type="checkbox"/> ideograph	<input type="checkbox"/> blotch	<input type="checkbox"/> raunchy
<input type="checkbox"/> lilting	<input type="checkbox"/> gibberish	<input type="checkbox"/> cluck	<input type="checkbox"/> smattering	<input type="checkbox"/> infuriate

국문초록

본 연구는 Laufer와 Hulstijn가 2001년에 제안한 과업 관여도 가설을 검증하고자 한다. 이 가설은 과업 관여도가 높을수록, 즉 학습자가 새로운 단어를 더 정교하게 처리할수록, 학습자가 그 단어들을 더 오랫동안 기억할 수 있다고 가정한다. 가설에 따르면 과업 관여도는 동기, 탐색, 평가라는 조작 가능한 세 구인으로 분석될 수 있다. 만약 유효성이 증명된다면, 이 가정은 연구자와 영어 교육자가 단어 학습 과업을 개발하고 평가하는 데 유용한 기준으로 기능할 수 있다. 지금까지 가설을 증명하려는 연구가 상당 수 수행되었지만, 그 결과가 일치하지 않는다는 점에서 가설을 검증하는 연구가 더 필요한 실정이다. 뿐만 아니라, 기존 과업 관여도 가설 연구는 대부분 대학생들을 대상으로 이루어졌고, 제 2언어 의미 이외에는 단어 지식의 다양한 양상을 분석하지 않았다. 또한 언어 학습에서 가장 중요한 변수 중 하나인 학습자 숙련도가 충분히 연구되지 않았다는 한계점이 있다.

본 연구의 참여자인 118 명 (57 명의 상급학습자와 61 명의 중급 학습자)의 한국 고등학교 3 학년 학습자는 각기 다른 과업 관여도를 지닌 세 과업 중 하나를 수행하고 즉시 사후 평가와 1 주일 후 이루어진 지연 사후 평가에 응하였다. 두 사후 평가는 모두 소극적 단어 학습 시험과 적극적 단어 학습 시험으로 구성되었다. 소극적 단어 학습 시험 점수와 적극적 단어 학습 시험 점수의 합을 분석하여 살펴본 결과, 단어 전반 학습 면에 있어서, 상급 학습자와 중급 학습자 모두 관여도 가설이 예측한 것과 거의 완벽하게 일치하는 학습 결과를 보였다. 즉시 사후 측정과 지연

사후 측정 모두에서 두 집단은 과업의 관여도 지수가 높을수록 더 높은 비율의 단어 학습을 보였다.

한편 과업 관여도의 소극적 단어 학습에 대한 영향은 적극적 단어 학습에 영향과는 다른 양상을 띠었다. 우선, 소극적 단어 학습 결과는 과업 관여도 가설에 대한 완벽한 증거가 될 수 있는 정도로 가설의 예측에 부합했다. 그러나 적극적 단어 학습 사후 측정 결과에서는 쓰기 학습이 상위 학습자에서만 통계적으로 유의미한 차이를 보였다는 것 외에는 관여도가 다른 과업들 사이에 큰 차이가 없었다. 즉 관여도 가설로는 적극적 단어 학습을 설명할 수 없었다. 따라서 관여도 가설의 설명력은 단어의 의미적인 영역에 제한될 뿐, 복잡한 단어 지식의 다양한 양상을 다 설명할 수 없을 수 있다. 적극적 단어 학습은 전이 적합성 이론과 같은 다른 이론들에 의해 더 잘 설명될 수 있거나, 혹은 처리의 깊이보다 학습자가 실제로 얼마나 많이 그리고 얼마나 오랫동안 새로운 단어를 처리하는가와 더 관련이 있을지 모른다. 결과에 근거하여, 본 연구는 형태기반 단어 수업에 대한 시사점과 이후 연구에 대한 제언을 결론부에 제시한다.

주요어: 관여도 가설, 단어 과업, 단어 학습 보존, 숙련도, 소극적 적극적 단어 학습

학 번: 2013-21372