교육학석사학위논문

Korean EFL High School Learners' Performance on Fill-in-the-blank Questions in Korean CSAT: Focusing on Amount of Textual Information, Blank Type, and Mock CSAT English Level

한국 EFL 고등학교 학습자의 대학수학능력시험 빈칸추론문항 수행: 지문 정보의 양, 빈칸의 종류, 모의고사 영어 등급을 중심으로

2019년 8월

서울대학교 대학원

외국어교육과 영어전공

김 지 은

Korean EFL High School Learners' Performance on Fill-in-the-blank Questions in Korean CSAT: Focusing on Amount of Textual Information, Blank Type, and Mock CSAT English Level

by

Ji Eun Kim

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

August 2019

Korean EFL High School Learners' Performance on Fill-in-the-blank Questions in Korean CSAT: Focusing on Amount of Textual Information, Blank Type, and Mock CSAT English Level

한국 EFL 고등학교 학습자의 대학수학능력시험 빈칸추론문항 수행:
지문 정보의 양, 빈칸의 종류, 모의고사 영어 등급을 중심으로

지도교수 이 병 민

이 논문을 교육학 석사 학위논문으로 제출함

2019년 6월

서울대학교 대학원

외국어교육과 영어전공

김 지 은

김지은의 석사학위논문을 인준함

2019년 8월

위 원 장 김 진 완

부위원장 소 영 순

위 원 이 병 민

Korean EFL High School Learners' Performance on Fill-in-the-blank Questions in Korean CSAT: Focusing on Amount of Textual Information, Blank Type, and Mock CSAT English Level

Approved by Thesis Committee:

Jin-Wan Kim, Committee Chair

Youngsoon So

Byungmin Lee

Abstract

Korean EFL High School Learners' Performance on Fill-in-the-blank Questions in Korean CSAT: Focusing on Amount of Textual Information, Blank Type, and Mock CSAT English Level

Ji Eun Kim

English Major, Dept. of Foreign Language Education

The Graduate School of Seoul National University

The fill-in-the-blank question is one of the most significant question types in Korean CSAT. The present study explored Korean EFL high school learners' performance on the fill-in-the-blank questions in Korean CSAT depending on the amount of textual information, blank type, and the English proficiency level (mock CSAT English level). 279 high school students with four English proficiency levels were provided with eighteen fill-in-the-blank questions from previous tests. The original texts were modified into four versions with different amount of textual information: a sentence with a blank (*Sentence-with-a-blank*), the *Sentence-with-a-blank* and its preceding and following sentences (*Left-and-right Passage*), the *Left-and-right Passag*e and the first and final sentences of the original passage (*First-and-final Passage*), and *Original Passage*. The questions also differed in the types of words fit in the blank: *One-word*, *Phrase*, and *Clause Blank*.

This study suggested learners' performance on the fill-in-the-blank questions with two major findings. First, reading the full text didn't provide the additional benefit, while reading the three core sentences led to the highest probability for correct answer for all learner groups. The lowest ability learners are predicted to perform the worst on the longest text, on which the highest ability learners displayed the highest probability for correct answer. Second, the learners with lower ability might experience stratified difficulty depending on the blank type while the learners with higher ability are less sensitive to the effects of blank type. The low-ability learners easily fail to read longer words required in the blank and fully comprehend the text to find the missing information whereas the high ability learners manage to do so.

To get the high score on the fill-in-the-blank question, not all the test-takers should read the whole passage; reading the three core sentences rather predict the highest probability for correct answer for all levels of learners and save time to read other questions. Although the probabilities for the correct answer on the full text and the three core sentences are not significantly different, it might be absurd to conclude that those two texts measure the same reading ability.

Keywords: fill-in-the-blank questions, test validity, multiple-choice questions, reading assessment, second language reading
Student number: 2017-27271

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

The present study explored Korean EFL high school learners' performance on the fill-in-the-blank questions in Korean CSAT depending on the amount of textual information, blank type, and the English proficiency level (mock CSAT English level). The current chapter discusses the background and purpose of the study in the beginning. To illustrate the concrete issue of the study, research questions and tentative hypotheses follow. Lastly, the outline of the thesis is overviewed, describing each chapter.

1.1 Purpose of the Study

The College Scholastic Ability Test (CSAT) has replaced the traditional standardized test since 1994, and it became the most significant high-stakes test with enormous washback effect as most of the Korean high school students take the test every year. The English section of the CSAT aims to assess the high school students' English ability which might be needed to take courses in college with the questions that match the content and the level of high school curriculum. Among the reading comprehension questions, the fill-in-the-blank question is one of the most significant questions. Many high school students find the question type more difficult than other types of questions. Due to the difficulty, the high scores are allocated on the fill-in-the-blank questions as the relatively large number of question appears on the test. Thus, the fill-in-the-blank question plays an important role in discriminating high-achieving learners

from the rest of the students in the CSAT.

Despite its significance, the questions in English section of CSAT including fill-in-the-blank questions bear weakness since they are all multiple-choice questions. In a multiple-choice test, test scores might be inflated or deflated since there are different ways of reaching the right answer (Nevo, 1989; Smith, 1982). Also, the multiple-choice test is susceptible to test-taking strategies (Anderson, Hierbert, & Wilkinson, 1985; Cohen, 1984, 2012) which were also adopted by ESL learners (Rupp, Ferne, & Choi, 2006). The learners regard answering a multiple-choice question as a problem-solving task instead of a reading comprehension task.

In addition, the fill-in-the-blank questions in the CSAT have been proven to be especially vulnerable to test-taking strategies. Test-taking strategies were often used on the question type according to the research on fill-in-the-blank questions. Oh (1999) and Nam (2015) explored the types of test-taking strategies used on the fill-in-the-blank questions in the CSAT with qualitative methods. In Oh's (1999) study, a Korean participant always read sentences with the blanks first rather than reading the text from the beginning to the end. Nam's (2015) recent study also confirmed Oh's (1999) study; the unusual reading patterns of checking the blank first was frequently found. In both of the studies, the participants were engaged in the partial reading or the mixed order reading to answer the fill-in-the-blank questions.

In reality, many students indeed read only parts of the text provided in the question. This partial reading behavior might stem from the insufficient reading fluency. Considering their reading fluency, the students do not have enough time to read the whole text to answer the questions. They have no alternatives but to read only the parts of the text, and they also try to guess the answer from the choices without reading the entire text.

The partial reading strategy is encouraged by instructions at many schools as well. In school, many teachers teach test-taking strategies to answer the fill-in-the-blank questions as a last resort for a high score. They believe students might reach the correct answer just by reading a few core sentences. In that way, students may also save time to answer other questions. Students are told to read a sentence with the blank first rather than to read the text from the beginning to the end. The test-taking skill is also promulgated in several programs by Educational Broadcasting System (EBS) and many private English institutes across the country.

As mentioned above, Korean high school students often read only parts of the text in responding to the fill-in-the-blank questions. Sometimes, it might be presumably meaningless to read the whole passage as long as they can find an appropriate choice among the five choices. When the students simply do not read the full text in a reading comprehension test, the test might not be able to appropriately measure what it intends to measure; when students read only parts

of the text, the fill-in-the-blank questions might measure different ability.

The current study attempts to examine whether the probability for correct answer differs depending on the amount of textual information that the participants get from the text. In other words, it attempts to confirm whether a fill-in-the-blank question could be answered without reading the entire passage, which is consider to be a common practice at schools and private institutes. Whether participants should read the whole text to get the correct answer or whether it is possible for them to reach the right answer just by reading a few core sentences as the teachers and students' belief will be investigated in the present study. In order to answer these questions, the present research set up several research questions.

1.2 Research Questions and Hypotheses

The research questions and tentative hypotheses of the present study are as follows:

1. How does the probability of getting correct answer change for each amount of textual information in the passage?

   Hypothesis: The probability for correct answer would be significantly different depending on the amount of textual information provided in the passage.

2. How does the probability of getting correct response change for each type of information required in the blank?

   Hypothesis: There would be some differences depending on the information required in the blank in terms of the probabilities for correct answer.

3. How does the probability of getting correct answer change for each English level?

   Hypothesis: Differences in the probabilities for correct answer among the English proficiency level groups will exist even using this type of questions.

4. What is the interaction effect of the amount of textual information and the English proficiency level groups?

   Hypothesis: There would be interaction effects of the amount of textual information and the English proficiency level groups. More advanced groups will presumably prefer longer texts to shorter ones, and their correct answer rate will be higher. Less advanced groups will prefer shorter passages, and their probability for correct answer will be higher in shorter text than in longer ones.

5. What is the interaction effect of blank type and English proficiency level groups?

   Hypothesis: There would be interaction effects of the blank type and the English proficiency level groups. Less advanced groups will be more sensitive to the blank type while more advanced will show similar probability for correct answer e across the blank type.

1.3 Organization of the thesis

This thesis includes five chapters: introduction, literature review, methodology, result and discussion, and conclusion. Introduction, the current chapter, offers the purpose of the study followed by specific research questions. The next chapter, literature review, introduces the theoretical background of the validity of second language reading test and the fill-in-the-blank questions in Korean CSAT that inspired this study. In the chapter, extensive studies on the characteristics of texts such as the amount of textual information and the blank type are subsequently reviewed. In Chapter 3, the research methodology is presented with participants, materials, procedure, and data analysis. Chapter 4 describes the results of the present study and discusses the result. The final chapter, Chapter 5, concludes the thesis with a summary of the major findings, their implications, the limitation of the study and the suggestion for future research.

Chapter 2. Literature Review

This chapter introduces the theoretical background and reviews the literature relevant to the present study. The process of reading comprehension and the validity of second language reading test are presented in the first section. In the second section, the previous studies on Korean CSAT and the fill-in-the-blank questions are reviewed. Subsequently, the literature on the characteristics of text such as the amount of textual information and the blank type follows in the last section.

2.1 Second Language Reading Test

2.1.1 Reading Comprehension

Recent studies on reading have a consensus that the reading process is inherently bottom-up. The reading process begins with lower-level processes and moves onto higher-level processes. The lower-level processes include word recognition, syntactic parsing, and semantic-proposition encoding (Grabe, 2009). Readers who fail to automatize the lower-level reading skills would have limited reading comprehension ability. The higher-level processes contain a text base of a reader's comprehension and a situation model of reader interpretation (Grabe, 2009).

Kintsch (1988) proposed a useful framework for explaining the reading comprehension process construction-integration (C-I) model. The C-I model includes a construction phase and an integration process. In the construction

phase, the linguistic input and the reader's knowledge construct a text base. In the integration process, the text base is integrated into a situation model. It is the final goal of reading comprehension which is built with the reader's experience and prior knowledge (van Dijk & Kintsch, 1983; Zwaan, 1999; Zwaan & Radvansky, 1998; Zwaan & Rapp, 2006). The C-I model implicates that the initial reading process is strictly bottom-up. Meanings of words are activated, propositions are forged, and inferences and elaborations are made. The reading process of ESL/EFL learners might be different from the reading process which the C-I model account for, especially in the test-taking context. In the present study, Korean EFL high school learners' reading process on the fill-in-the-blank questions is discussed in comparison with the reading process presented in Kintsch's C-I model.

2.1.2 Validity of Second Language Reading Test

As reading is a multi-faceted skill which is hard to achieve even in a first language (Kintsch & Kintsch, 2005; Rapp & van den Broek, 2005; Rayner, Pollatsek, Ashby, & Clifton, 2012), assessing reading comprehension is never an easy task. Reading comprehension test like other language tests involves multiple qualities such as reliability, construct validity, authenticity, interactiveness, impact, and practicality. Among those qualities, validity is an essential quality to be considered in reading comprehension tests as well as in other language test. According to Messick (1989, 1995, 1998), validity is the

process in which test developers collect evidence for a particular interpretation of test results. It is not merely what a test tries to measure but an attempt to understand what a test score means.

However, the validity might be threatened by the construct underrepresentation and the construct-irrelevant variance (Akbari, 2012). The construct underrepresentation accounts for the adequacy of the sample and the chance that the sample is less than the performance in the target language domain. The construct-irrelevant variance is concerned with the involvement of some irrelevant factors that might affect the test score. The unwanted elements hold the risk that the test result may lead to an incorrect interpretation. For example, a multiple-choice test might be susceptible to a construct-irrelevant variance like the use of test-taking strategy.

The discussion over the validity of reading comprehension tests with multiple-choice questions was vigorous. The problem of multiple-choice lies on the doubt that these tests may trigger the test-taking strategies. Since the participants use test-taking strategies which only direct to the correct answer (Anderson et al., 1985; Cohen, 1984), they do not read or comprehend the entire reading passage in responding to the multiple-choice test. The participants exploit diverse strategies and mental resources to get the correct answers. That way, the test scores might be inflated or deflated (Smith, 1982), and the interpretation of the test scores cannot generalize the tasks in the target language

domain. More interestingly, according to the literature (*e.g.*, Cohen, 1984; Katz, Lautenschlager, Blackburn, & Harris, 1990; Daneman & Hannon, 2001), the participant who did not read the passage still performed well above the chance level. Studies in ESL contexts reported that the participants employ a variety of test-taking strategies. Rupp et al. (2006) purport that ESL learners in their study answered multiple-choice questions not as a comprehension task but as a problem-solving task.

Furthermore, multiple-choice questions might bring a washback effect in the classroom (Johnston, 1987; Mayher & Brause, 1986). The instructions on multiple-choice tests can limit readers' interpretation of the text into a single correct one rather than allow creative interpretations of the text. Students are encouraged to prepare for better performance on the standardized achievement tests, restricting the students' diverse interpretation of the passage (Anderson, Hiebert, & Wilkinson, 1985; Frederiksen, 1984). Also, the instruction given for the multiple-choice tests forces attention to the surface forms and does not guide to in-depth reading (Bussis & Chittenden, 1987).

All in all, the validity issue of reading comprehension test suggests the necessity of including information about how participants process tasks, and of connecting different kinds of information on test content and performance. That is to say, to validate the reading comprehension test, not only the linguistic components but also what happens during the test should be taken into

consideration.

2.2 Fill-in-the-blank Questions in Korean CSAT

2.2.1 English Section of Korean CSAT

Since 1994, the traditional standardized test has been replaced by the College Scholastic Ability Test (CSAT) (Kim, 2015; Kwon, 2015) which became the most significant high-stakes test for Korean high school students with substantial washback effect (Choi, 2008). According to the test-developer, Korea Institute for Curriculum and Evaluation, the goals of the CSAT are to measure the scholastic ability which might be needed in the college education and to provide questions that match the content and the level of the high school curriculum. In the English section of the CSAT, students' English ability is measured by forty-five multiple-choice questions. The forty-five multiple-choice questions consist of seventeen listening questions and twenty-eight reading comprehension questions. Since the listening questions take approximately twenty-five minutes, participants have less than forty-five minutes to answer the twenty-eight reading comprehension questions.

The English section of CSAT was welcomed in the early years of adoption because the test was believed to tap creative thinking over memorization of simple knowledge. The washback effect of the CSAT had been considered positive in classroom learning and teaching. Song (1998) argued the students who took the CSAT were less dependent on bottom-up reading

strategies than the students who entered university before the CSAT implementation were. The test had also been believed to provide reliable test results (Kim, 2015).

However, the original purpose of the English section of CSAT was tainted along with the numerous policies like the policy that forces to involve materials developed by Education Broadcasting System, the one percent of people with a perfect score policy, and easy CSAT English policies. Some researchers criticized that the CSAT failed to achieve its main goals: assessing the general scholastic ability to take a university course as originally intended (Kim, 2015) and providing questions that match the content and the level of high school curriculum (Kim, 2001; Kim & Ma, 2012). Three major criticisms on the English section of CSAT are as follows.

First, the CSAT does not reflect the high school curriculum while one of the goals of the test is to provide questions that match the curriculum. The texts presented in the test do not match with high school text books in a few ways. The texts given in the test are more challenging than the texts provided in the high school textbooks. In Kim and Ma's (2012) study, the 2012 CSAT was higher in Flesch-Kincaid Reading Grade Level and lower in Reading Ease Score than the seven high school textbooks in the study. The CSAT also seemed to obtain more academic words than the textbook counterparts. Furthermore, the types of questions in CSAT is strikingly different from those from the high

school textbooks. Kim (2001) argued there was a significant difference in the listening and reading question types between the 1999-2001 CSAT and the five textbooks examined in the research. She continued that neither of the CSAT nor the textbooks faithfully reflected the high school curriculum. These results imply that studying English by high school textbooks is not enough to get a high score in the CSAT.

Second, the quality of distractor which is closely related to the general test-taking skill is considered to be one of the most critical attributes of the test difficulty in the CSAT. The item difficulty predictors of the CSAT were investigated by Chon and Shin (2010). They proposed that the plausibility of incorrect distractors was one of the most significant predictors of difficulty. It was also supported by Chang's (2004) research on a model of predicting item difficulty of the CSAT. In her final model, the attractiveness of distractor was involved as one of the most influential factors in predicting the difficulty. However, the quality of distractor accounts for general test-taking skills rather than reading comprehension skill (Ozuru, Rowe, O'Reilly, & McNamara, 2008).

Third, the washback effects of the CSAT also led the classroom to focus on 'not making mistake training' rather than actual learning (Kim, 2015). The use of test-taking strategies and the instruction of test-taking strategy might be induced. Song (1998) revealed that the high and low ability group might have a difference in test-taking strategy use in the CSAT. The high ability group were

more accustomed to using test-taking strategies and used various strategies than the low ability group (Haam, 2006). Also, Choi (2008) contended that the majority of Korean high school students have to learn test-taking strategies instead of improving their English ability due to the tremendous washback effects of the CSAT. The washback effects of the CSAT deprive the students of opportunities to acquire productive language skill.

2.2.2 Fill-in-the-blank Questions

In the CSAT, the fill-in-the-blank question is one of the most important types of question for the following reasons. First, the fill-in-the-blank question is the most difficult question type for students. Jin and Park (2004) revealed fourteen significant variables out of sixty-four variables that affect item difficulty in the CSAT. One of the significant variables was *inferring the parenthesis (phrase)* which is the fill-in-the-blank question with phrase blank. Another work done by the same authors confirmed that the fill-in-the-blank is one of the sixteen main factors affecting the difficulty of the CSAT (Jin & Park, 2005). The difficulty of the questions might be due to the text difficulty; texts provided in the questions are more demanding than texts in other questions. The required task of filling in the blank on the question type also seems to be challenging. One of the sources of the difficulty might be due to the lack of chances to practice those question types since the fill-in-the-blank questions rarely appear in the high school textbook (Kim, 2001).

Second, the importance of the fill-in-the-blank questions is also supported by the allocated score and the number of questions. The scores allocated in the fill-in-the-blank questions are higher than the allocated scores of other questions. For instance, three points are usually allocated for most of the fill-in-the-blank questions whereas two points are allocated to other questions. Also, the question type takes up fourteen percent of the reading comprehension question, which is the relatively large number compared to other questions. Therefore, the fill-in-the-blank questions in the CSAT play a significant role in discriminating high ability learners from the rest of the learners.

Given the high difficulty and importance of the fill-in-the-blank questions, many test-takers implement test-taking strategies to answer the question type. Oh (1999) observed test-taking strategies on the fill-in-the-blank questions with two participants: a Korean university student and an English-speaking professor. The Korean participant wrote down her thoughts while responding to the full version of seventeen fill-in-the-blank questions. She always read the sentences with the blank first instead of reading from the beginning of the passage. The English-speaking participant was provided with only a sentence with the blank from the same fill-in-the-blank questions. He answered ten questions correctly out of seventeen items just by reading one sentence. Oh argued that the fill-in-the-blank questions induce reading process that are rarely observed in the normal reading process (Alderson, Clapham, &

Wall, 1995). Therefore, this type of reading comprehension question prohibits the normal reading process (Hughes, 1989).

Nam (2015) also explored the test-taking strategies used on the fill-in-the-blank questions with eight participants who reported their test-taking process with the think-aloud protocol. The participants did not read the passage from the beginning except for one test item and did not always read the whole passage to answer the questions as in Oh (1999). Also, Kim and Chon (2014) compared the test-taking strategies used on three types of reading comprehension questions in CSAT. They concluded that more test-wiseness strategies are used on the fill-in-the-blank questions than on other questions in CSAT. For the fill-in-the-blank questions, test-wiseness strategies and test-management strategies are most often employed. The concept of test-wiseness strategies and test-management strategies are more discussed in the following section.

2.2.3 Amount of Textual Information in Fill-in-the-blank Questions

The effects of text length on reading comprehension (Yi, 2013) which is a seemingly similar concept to the amount of textual information were controversial. Some researchers (*e.g.*, Rothkopf, 1965, Cha, 1995; Commander & Stanwy, 1997; Chujo & Utiyama, 2005; Cater, Walker, O'Brien, & Hough, 2017) asserted that the longer the text, the better the reading comprehension. Some studies (*e.g.*, Mehrpour & Riazi, 2004; Beach, 2008) failed to confirm the benefit of long text. On the other hand, other researchers' argument (*e.g.*,

Rothkopf & Billington, 1983; Andreassen & Bråten, 2010) showed the rather opposite result; they asserted shorter text is beneficial for reading comprehension. Andreassen and Bråten (2010) disputed that the longer the text in a multiple-choice test, the use of working memory increases. Nevertheless, the text length controlled in the previous studies might be fundamentally different from the amount of textual information manipulated in the present study. While the short text was considered to be about 200 words in previous studies, the longest text in the present study holds about 150 words.

The amount of textual information adjusted in the present study is rather concerned with test-taking strategies. The test-taking strategies are the strategies that learners consciously select while they are taking tests (Ghafournia & Afghari, 2013). Like learning strategies, the test-taking strategies could be effective or ineffective since the effects of used strategies rely heavily on whether the learners use the strategies appropriately in specific tasks. When appropriate strategies are used, low proficient learners might end up with the shortcut to the answer without completing any linguistic task. In other words, participants might come up with test-wiseness to avoid using their real linguistic knowledge to answer the test items. Test-wiseness is the ability to utilize test-taking strategies to choose the correct answer on multiple-choice items (Allan, 1992). The test-wiseness enables participants to respond without necessarily reading and understanding the whole passage on the test. As Allan indicated,

test-taking strategies could be the source of the construct invalidity. Given that test-wiseness strategies influence reading test result, the test scores might not account for the participants' reading comprehension skill.

To rule out the confusion of terminology, Cohen's (2012) three types of test-taking strategies are described here: language learner strategies, test-management strategies, and test-wiseness strategies. The language learner strategies are activated for the basic skills of listening, speaking reading, and writing. Test-management strategies encourage participants to answer the test items meaningfully with returning to the question, comparing alternatives in multiple-choice items and crosschecking the options with the text to make sure the option seems correct. Test-wiseness strategies which are the primary interest of the present study are the strategies for implementing the knowledge of the characteristics of the test format itself and other peripheral information without using the linguistic and cognitive process that test developers might expect participants to utilize. Eliminating the alternative which seems wrong by common sense is the example of the test-wiseness strategies.

In his seminal study on test-taking strategy, Cohen (1984) introduced five unpublished course papers on test-taking strategies conducted by his pupils. In experiment 1 and 2, only a quarter of the participants read the entire passage to answer multiple-choice tests. In the other two experiments introduced by Cohen (1984), the participants read the questions before they read the text in

multiple-choice questions. They stopped reading the alternatives as soon as they found the correct answer. Farr, Pritchard, and Smitten (1990) also confirmed the result; less than half of the participants read the entire passages before they read the questions. They proclaimed that answering for questions had a stronger emphasis over reading strategies. When parts of the passage were given in the fifth experiment of Cohen (1984), 49 percent of advanced and 41 percent of intermediate participants got correct answers. Considering that the chance level is 25 percent in the multiple-choice questions with four alternatives, the probability for the correct responses was quite high. The study offers that "respondents get items wrong for the right reasons or right for the wrong reasons" (Cohen, 1984, p.71).

According to subsequent studies on test-wiseness strategies, readers might not start reading the passage from the beginning in the reading comprehension tests. "Looking both at words and phrases immediately preceding and following a blank, as well as at the extended context for clues as to how to fill it in" (Cohen, 2012, p.100) is the typical test-wiseness strategies dealing with cloze tests. Similar strategies were introduced by Oh (1999) and Haam (2006) as the test-wiseness strategies dealing with the fill-in-the-blank questions. Reading the immediately preceding and following a blank is assumed to be effective test-wiseness strategies on fill-in-the-blank questions.

2.2.4 Blank Type in Fill-in-the-blank Questions

The blank type, as well as the amount of textual information, is an interest of the present study which might affect the learners' performance on fill-in-the-blank questions. The fill-in-the-blank questions require one-word, phrase, or clause in the blank. Although the blank type is a scarcely researched area, the categorization of passages with blank types is found in the literature. Haam (2006) might believe the *filling the blank with adequate conjunctions* and the *filling the blank with adequate paragraph* are different test items as different test-wiseness strategies are suggested in the study. On their research on item difficulty, Jin and Park (2004) set *inferring the parenthesis (words), inferring the parenthesis (phrases)*, and *inferring the parenthesis (sentences)* as different item types. Therefore, the classification of blank types into one-word, phrase, and clause is intuitively plausible and matches the previous study.

In addition, the fill-in-the-blank question which is completed by a phrase might be more challenging than the question which requires a single word or clause in the blank. Jin and Park (2004) showed that *inferring the parenthesis (phrases)* is a significant predictor variable on item difficulty while the *inferring the parenthesis (words)* and *inferring the parenthesis (sentences)* were removed since they were not significant factors on item difficulty.

The previous research on the fill-in-the-blank question was mainly focused on types of test-wiseness strategies employed in the CSAT; most of the

learners read the sentence including the blank first and did not finish reading. Since few studies revealed whether the strategic reading behavior contributes to the probability for correct answer, the current study aims to investigate the effect of the amount of textual information and blank type on learners' performance on fill-in-the-blank questions

Chapter 3. Methodology

The overall research was designed to explore which predictor variable (*i.e.*, amount of textual information, blank type, and English proficiency level) affects Korean EFL high school students' probability of getting a the fill-in-the-blank question correct. This chapter introduces the method employed in the study. First, the overall research design, participants, and materials used in the present study are introduced. The materials contain the main task and the questionnaire. Afterward, the procedure of the experiment along with the pilot study follows. In the last section, the statistical analysis method, mainly the multilevel logistic regression model, is presented.

3.1 Participants

Two hundred seventy-nine eleventh-grade Korean high school students in four coeducational high schools in Daejeon, Jeon-ra, and Gyeong-gi provinces in South Korea voluntarily participated in the study. Eleventh-grade students were chosen for the participants because most of them will take the College Scholastic Ability Test (CSAT) in the following year. Since they were already preparing for the CSAT, they have enough English proficiency to answer the test items presented in the study. Moreover, most of them are not likely to be familiar with the previous questions which were used in the present study. To minimize the familiarity effects of the reading passages used in the study, the English teachers of the participating schools confirmed that the students had not studied the texts

before the experiment. The data which reported the topic familiarity by the questionnaire were excluded from the final analysis.

To scrutinize the effects of the English proficiency level on the students' performance, the participants were divided into four groups based on the English proficiency levels. All the participants took the same mock version of CSAT for eleventh-grade students developed by the Korean Institute for Curriculum and Evaluation (KICE) one month before the experiment. For the present study, the Level 1 group is ranged from 90 to 100; the Level 2 group is ranged from 80 to 89; the Level 3 group is ranged from 70 to 79; and the Level 4 group is ranged from 0 to 70 (see Table 3.1).

Table 3.1
Description of the Participants based on English Proficiency level

English level (present study)	Mock CSAT score	Number of participants
Level 1	90-100	55
Level 2	80-89	49
Level 3	70-79	56
Level 4	0-70	119
Total		279

3.3 Materials

3.3.1 Main Task

Eighteen fill-in-the-blank questions from both the previous CSATs and the mock versions of CSAT from 2009 to 2013 were randomly selected for the present study. Table 3.2 displays the eighteen items selected for this study and other relevant information. The main reason to choose somewhat old items from the previous CSAT is to control the familiarity factor in the study. Since the selected items are drawn from 2009 to 2013, the participants are less likely to have come across the reading passages used in the study.

Table 3.2
Features of the Selected Eighteen Items for the Study

Item number	Year	Word count per passage	Lexile measure	Flesch-Kincaid Grade level
1	2009	127	900L - 1000L	10.2
2	2009	115	1000L - 1100L	9.2
3	2010	102	700L - 800L	6.9
4	2010	122	1000L - 1100L	11
5	2010	137	900L - 1000L	9.6
6	2010	143	1200L - 1300L	13.1
7	2011	126	1100L - 1200L	10.8
8	2011	136	800L - 900L	9.7
9	2011	126	1300L - 1400L	14.7
10	2011	125	1000L - 1100L	7.7
11	2012	136	1100L - 1200L	10.8
12	2012	141	1100L - 1200L	10.1
13	2012	135	1100L - 1200L	7.9
14	2012	119	1000L - 1100L	8.9
15	2013	123	1100L - 1200L	11.1
16	2013	150	600L - 700L	5.4
17	2013	137	600L - 700L	6.6
18	2013	147	1200L - 1300L	10.3
	Mean	130	983L-1083L	9.6

The reading passages from the original test items are modified into four different versions depending on the amount of textual information. The amount of textual information was manipulated to reflect those typical strategies which are known as effective rather than modified randomly since using the appropriate test-wiseness strategies is more effective on the test-taking performance than just using many test-wiseness strategies (Dolly & Willliam, 1986; Ghafournia & Afghari, 2013, Nevo, 1989). According to the previous studies on test-wiseness strategies (*e.g.*, Haam, 2006; Oh, 1999), reading immediately predceding and following a blank is the frequently employed test-wiseness strategies.

(a) An *Original passage* is the original test item from the previous tests without any modification. The *Original passage* is offered as a control condition.

Example:

When we behave irrationally, our behavior usually seems reasonable to us. When challenged, the mind says (to itself), "Why are these people giving me a hard time? I'm just doing what makes sense. Any reasonable person would see that!" In short, we naturally think that our thinking is fully justified. As far as we can tell, we are only doing what is right and proper and reasonable. Any fleeting thoughts suggesting that we might be at fault typically are [_____] by more powerful self-justifying thoughts: "I don't mean any harm. I'm just! I'm fair! It's the others who are wrong!" It is important to recognize this nature of the human mind as its natural state. In other words, humans don't have to learn self-justifying, self-serving, self-deceptive thinking and behavior. These patterns are innate in every one of us.

a) spread b) unveiled c) fortified d) overcome e) authorized

(b) A *Sentence-with-a-blank* has a single sentence with one blank.

Example:

Any fleeting thoughts suggesting that we might be at fault typically are [_____] by more powerful self-justifying thoughts: "I don't mean any harm. I'm just! I'm fair! It's the others who are wrong!"

a) spread b) unveiled c) fortified d) overcome e) authorized

(c) A *Left-and-right Passage* contains one sentence with the blank as in the *Sentence-with-a-blank* condition, and its neighboring left and right sentences are also provided. The left sentence is the preceding sentence of the *Sentence-with-a-blank*, and the right sentence is the following sentence of the *Sentence-with-a-blank*.

Example:

As far as we can tell, we are only doing what is right and proper and reasonable. Any fleeting thoughts suggesting that we might be at fault typically are [_____] by more powerful self-justifying thoughts: "I don't mean any harm. I'm just! I'm fair! It's the others who are wrong!" It is important to recognize this nature of the human mind as its natural state.

a) spread b) unveiled c) fortified d) overcome e) authorized

(d) In a *First-and-final Passage*, the three sentences as in the *Left-and-right Passage* and the first and final sentences of the original passages are provided. Other sentences from the original passage are removed and left as a blank space. To avoid any influence from the visual oddness of the *First-and-final Passage* condition, the participants were told that some sentences are intentionally deleted. Some of the first and the final sentences of the passage overlap with sentences of the *Left-and-right*

Passage version but the *First-and-final Passage* was one or two sentences longer than the *Left-and-right Passage* in every case.

Example:

When we behave irrationally, our behavior usually seems reasonable to us.

Any reasonable person would see that!" In short, we naturally think that our thinking is fully justified. As far as we can tell, we are only doing what is right and proper and reasonable. Any fleeting thoughts suggesting that we might be at fault typically are [_____] by more powerful self-justifying thoughts: "I don't mean any harm. I'm just! I'm fair! It's the others who are wrong!" It is important to recognize this nature of the human mind as its natural state.

These patterns are innate in every one of us. (The blank space is intentionally left as blank so that the readers know the sentences are not directly connected).

a) spread b) unveiled c) fortified d) overcome e) authorized

In addition, the blank types are categorized into three groups: *One-word Blank*, *Phrase Blank*, and *Clause Blank*. Note that the filling in the blank with conjunctions is no longer exploited in the recent CSAT, which is why it is excluded from the present study. The followings are the examples of the three question types.

(a) A *One-word Blank* question requires a test-takers to fill in a single word to complete the blank-sentence.

Example:

Magicians are honest deceivers. To investigate the secret used by magicians to fool their audiences, Jastrow worked with two great illusionists. He invited these performers to his laboratory and had them

participate in a range of tests measuring their speed of movement and accuracy of finger motion. But Jastrow's results revealed little out of the ordinary. He demonstrated magic has little to do with fast movements. Instead, magicians use a range of [_____] weapons to fool their audiences. The technique of suggestion, which captures people's minds, plays a key role in the process. In the same way that people can be made to believe that they once went on a non-existent trip in a hot-air balloon, so magicians have to be able to manipulate people's perception of performance.

a) Ethical b) political c) physical d) economic e) psychological

(b) A *Phrase Blank* question requires a phrase to fill in the blank. A *Phrase Blank* holds a noun phrase or a verb phrase. The correct responses in the *Phrase Blank* of the current study are approximately 5.6 words on average. Among the six *Phrase Blank* items presented in the study, two of them have noun phrases and the other four have verb phrases.

Example:

About three percent of the weight of sea water is accounted for by salt. The salt content of the ocean is being continually added to by runoff from the land, but it is not increasing. Yet salt cannot leave the sea by evaporation because the water vapor leaves it behind. So for many years it was a mystery as to why the sea should [_____], given the rate of runoff, and particularly if the world was supposed to be hundreds of millions of years old. The riddle was not solved until the 1970s, when the deep ocean openings were discovered. Sea water is swallowed up by these cracks in the ocean bed; when it reemerges, millions of years later, as steam from volcanoes, the salt has been filtered out of it on its passage through the rocks.

a) be so deep and large b) create so much runoff

c) have so little salt in it d) have so many volcanoes

e) keep its temperature so steady

(c) A *Clause Blank* question can be completed with a clause. Average number of words fit in the *Clause Blank* questions in the current study a 9.3 words.

Example:

When we have made an error, as for example in adding up a column of figures, we have a tendency to repeat it again and again. This phenomenon is known as the persistent error. The same thing happens when we try to solve a problem; each time our thoughts take a certain course, that course is more likely to be followed the next time. The reason that we keep making the same error repeatedly is that associations form between the ideas in the chain of thoughts and become firmer each time they are used, until finally the connections are so well established that [_____]. Thus, once we have adopted an unprofitable line of thought, it is harder to adopt a profitable line.

a) the chain is very difficult to break

b) persistent problems are solved automatically

c) the ideas lose their associations with one another

d) those connections become weaker as time goes by

e) the phenomenon will lead to a profitable line of thought

To rule out the effect of item difficulty, the selected eighteen questions are counterbalanced on four test formats (*i.e.*, Format A, B, C, and D), and every participant answered one of the four formats (see Table 3.3).

Table 3.3
Features of Texts in Four Test Formats Used in the Study

Item No. (Blank type)	Amount of textual information			
	Format A	Format B	Format C	Format D
1 (*One-word*)	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*
2 (*Clause*)	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*
3 (*One-word*)	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*
4 (*Clause*)	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*
5 (*Phrase*)	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*
6 (*One-word*)	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*
7 (*Clause*)	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*
8 (*One-word*)	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*
9 (*Clause*)	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*
10 (*Phrase*)	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*
11 (*One-word*)	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*
12 (*Phrase*)	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*
13 (*Clause*)	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*
14 (*Phrase*)	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*
15 (*Phrase*)	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*
16 (*Phrase*)	*Original Passage*	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*
17 (*One-word*)	*Sentence-with-a-blank*	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*
18 (*Clause*)	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*	*Sentence-with-a-blank*

3.3.2 Questionnaire

The questionnaire was implemented after the main task to investigate the topic familiarity effect and to explore the learners' test-wiseness behaviors. The participants were asked to report whether they already read the texts prior to this study. All data reporting topic familiarity were excluded from the final analysis. Out of 5022 items (279 participants*18 items), 39 items were reported to be familiar to the participants.

Furthermore, the participants reported their general test-wiseness behaviors on the mock version of CSAT and particularly on the fill-in-the-blank questions. Although the main interest of the present study is to investigate the predictor variables on the probability of getting a question correct, the questionnaire might provide their story of test-wiseness behaviors on the fill-in-the-blank questions. How frequently the test-wiseness strategies of interest is employed could be investigated by the questionnaire. The questionnaire incorporated six questions about test-wiseness strategies. The first four of them could be answered in a five-point Likert scale (1=strongly disagree; 5=strongly agree), and the others were multiple-choice questions. All the questions were written in Korean. The six questions are as follows:

Q1. I read the whole passage in each item when I take an English reading comprehension section of CSAT (1=strongly disagree; 5=strongly agree).

Q2. I read the whole passage when I answer the fill-in-the-blank questions in the CSAT (1=strongly disagree; 5=strongly agree).

Q3. I read the questions and alternatives before I read the passage when I take an English reading comprehension section of CSAT (1=strongly disagree; 5=strongly agree).

Q4. I read the questions and alternatives before I read the passage when I answer the fill-in-the-blank questions in the CSAT (1=strongly disagree; 5=strongly agree).

Q5. I read [the first sentence of the passage, the final sentence of the passage, the preceding sentence of the blank-sentence, the blank-sentence, the following sentence of the blank-sentence] first when I answer the fill-in-the-blank questions in the CSAT. (The words in the square brackets are the alternative responses.)

Q6. I don't read the passage in the fill-in-the-blank questions from the beginning because of [the time pressure, the item difficulty, the instruction].

3.4 Procedures

The task materials were piloted before the main study to verify the difficulty of the test items, to predict the topic familiarity effect, and to test the appropriateness of the questionnaire.

3.4.1 Pilot Study

The participants of the pilot study were nine university freshmen who took the CSAT last year. They are considered to have similar cognitive and linguistic ability to the participants of the main study. Also, the maximum effect of topic familiarity could be predicted from them because they studied for the CSAT for the longer time. Based on their CSAT scores from the previous year, there were four Level 1 participants, four Level 2 participants, and one Level 3

participant.

According to the result, the longer text did not lead to the higher possibility of correct response (see Table 3.4). After the pilot study, two test items were replaced with other items in the main study because none of the participants answered correctly on the two items. Since the participants of the main study were two years younger than the participants of the pilot study, the two items seem to be too difficult to answer. Considering the years that they have studied for the CSAT, the topic familiarity effect was not noticeable. They reported that 3.4 items on average were familiar out of 18 items. The questionnaire was confirmed to be clear and understandable to answer.

Table 3.4
Rate of Correct Answer Depending on Amount of Textual Information (Pilot Study)

	Sentence-with-a-blank	*Left-and-right Passage*	*First-and-final Passage*	*Original Passage*
Level 1	40	63.75	60	55
Level 2	35	60	64.16	60
Level 3	0	40	50	0

3.4.2 Main Study

The main study is comprised of the main task and the questionnaire. Before the main study, the researcher explained the purpose of the study to the participants and made sure they voluntarily participate in the research. The

participants answered the main task first and responded to the questionnaire. In the main task, the participants answered the eighteen fill-in-the-blank questions with four textual information conditions and three blank types, as mentioned in the previous sections. Except for the predictor variables of interest, other test-taking conditions were equally controlled.

One minute and thirty seconds were given for each test item regardless of the amount of textual information. Excluding the listening question session, the participants have less than forty-five minutes to answer the twenty-eight reading comprehension questions. Thus, one minute and thirty seconds were calculated to reflect the average amount of time to solve each test item. The underlying reason for the time constraint is that the limited time in the standardized test might induce certain test-taking behaviors like the partial reading of the text. Taking Korean EFL high school students' reading fluency into account, the time constraint given in the CSAT might be responsible for the unique reading process in a test-taking setting. The researcher used a stopwatch and told the participants to answer the next test item one by one. The participants who already answered a test item within the time limit were asked to wait for others to finish. One question is printed on a single page so that participants can move onto the next question only after the researcher's sign.

After finishing the main task, the participants answered the questionnaire. In the questionnaire, the participants reported the topic-familiarity and their

general employment of the test-wiseness strategies in the real testing context. At the end of the experiment, the original test items, answer key, and the possible explanation were provided to the participants for the educational purpose. Small gifts were also given as an appreciation for participation.

3.5 Data analysis

The main task was analyzed by means of multilevel logistic regression since each individual participant answered eighteen questions which were combinations of the predictor variables (*i.e.*, amount of textual information and blank type). Since the responses were nested within individual students, the data were hierarchically structured. To examine hierarchically nested data within groups, the multilevel model, also known as mixed effect model or hierarchical linear model, was developed (Palmer, Graham, White, & Hansen, 1998) and has been recognized as a useful tool of statistical analysis in many sub-fields of linguistics (Eager & Roy, 2017).

In the present study, the multilevel modeling was appropriate to investigate the fixed effects which account for predictor variables (*i.e.*, amount of textual information, blank type, and English proficiency level) and the random effect of individual participants. To assure the independent observation of individuals, the intraclass correlation was calculated, using the formula given by Heck, Thomas, and Tabata (2012):

$$\rho = \frac{\sigma^2}{\sigma^2 + 3.29}$$

The intraclass correlation (ICC) was 0.115, suggesting that about 11.5% of the variability exists between individuals. An ICC value of .10 might be considered as a medium effect (LeBreton & Senter, 2008). Note that ICC merely reflects what the world is like in terms of the responses of the individuals are similar over time.

In addition, the logistic regression was useful to determine how independent variables (*i.e.*, amount of textual information, blank type, and English proficiency level) influence the binary dependent variable (*i.e.*, correct or incorrect). The logistic regression which is similar to linear regression can be adopted when the dependent variable is binary or dichotomous (Greenhouse, Bromberg, & Fromm, 1995). The logistic regression aims to find the best fitting model to describe the relationship between independent or predictor variables and a dependent variable (Hosmer & Lemeshow, 2000). The logistic regression models in the study were fit to explain the data.

Since the independent variables in the current study are all categorical variables, one category within a variable was set as a reference. The odds ratios in logistic regression are expressed in comparison with the reference category. The reference category refers to a baseline against which other categories are compared. For the amount of textual information variable, the *Original Passage* was chosen as a reference since it is the control condition of fill-in-the-blank questions which appears in real test-taking context. For the question types in the

blank variable, the *One-word Blank* was set as a baseline category, for it is predicted to be the easiest. For the English proficiency level variable, the Level 4 with the lowest English level was selected as a baseline because it had the most participants.

In the analysis of the questionnaire, one-way ANOVA and correlation were practiced to investigate the learner's test-taking behavior. One-way ANOVA was conducted to investigate whether there is any relationship between the test-taking behavior and the English proficiency level since the first four questions could be answered with a five-point Likert scale. Also, the correlation between reading the full text from the beginning to the end and the total score of main task was calculated. All statistical analyses were done with IBM SPSS Statistics 25.

Chapter 4. Results & Discussion

This chapter presents the results of the statistical analyses to answer the research questions presented in Chapter 1. First, to verify the predictor variables (*i.e.*, amount of textual information, blank type, and English proficiency level) on the probability of getting correct answer in the fill-in-the-blank questions, all the predictor variables and interactions were entered in the first multilevel logistic model (Model 1). Second, the final model of multilevel logistic (Model 2) which includes only significant predictor variables and interactions is presented. The initial hypotheses are discussed along with the findings obtained from the statistical analyses. The questionnaire was analyzed with one-way ANOVA and correlation.

4.1 Main Task

To explore the predictor variables (*i.e.*, amount of textual information, blank type, and English proficiency level) on the probability of getting correct answer in the fill-in-the-blank questions, 279 high school students answered eighteen questions each. The total responses included in the present study was 4983, excluding the 39 responses reporting the topic familiarity. The texts presented in the questions were modified into four text conditions with different amounts of textual information: *Sentence-with-a-blank*, *Left-and-right Passage*, *First-and-final Passage*, and *Original Passage*. The fill-in-the-blank questions

were also categorized into three different types of words fit in the blank: *One-word Blank*, *Phrase Blank*, and *Clause Blank*. Finally, the participants were divided into four levels by the English proficiency level, from the lowest Level 4 to the highest Level 1. There were total forty-eight levels (4×3×4) within three categorical variables, and the sample size of each level was more than enough to practice statistical analysis. Figure 4.1 and Table 4.1 shows the rate of correct answer for each combination of the amount of textual information, blank type, and the English proficiency level.

Despite the small amount of information given on the *Sentence-with-a-blank* and the *Left-and-right Passage*, the rate of the correct in those text conditions was well above the chance level. This confirms the literature on test-wiseness strategies. When only parts or none of the texts were given to the participants, the participants still performed above the chance level (Cohen, 1984; Katz et al., 1990; Daneman & Hannon, 2001). Moreover, according to the previous studies on the fill-in-the-blank questions in CSAT, the participants mostly read the sentence with a blank first rather than read the text from the beginning to the end (*e.g.*, Oh, 1999; Nam; 2015). When parts of the texts were given in the present study, the performance of the participants was above the chance level of 20%. Surprisingly, reading more information did not help all the learners to get a higher score than reading only one sentence as in *Sentence-with-a-blank* or three sentences as in *Left-and-right Passage*.

Also, the *Phrase Blank* questions and the *Clause Blank* questions were a lot more challenging than the *One-word Blank* questions. The result confirmed the previous study on item difficulty in CSAT. In Jin and Park (2004), *inferring the parenthesis (phrase)* was revealed as a significant predictor variable on item difficulty in CSAT while *inferring the parenthesis (words)* and *inferring the parenthesis (sentence)* were insignificant.

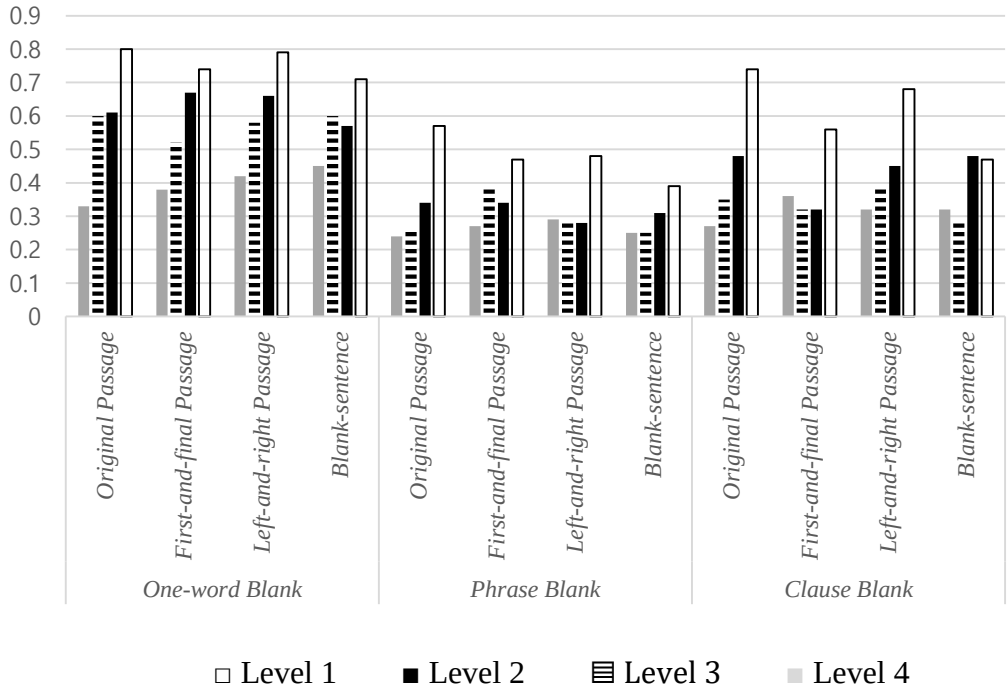

Figure 4.1 Rate of correct answer for amount of textual information, blank type, and English proficiency level

Table 4.1

Rate of Correct Answer for Amount of Textual Information, Blank Type, and English Proficiency Level

Blank type	*Amount of textual information*	Level 1		Level 2		Level 3		Level 4	
		N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
One-word Blank	*Original Passage*	69	0.80 (0.405)	71	0.61 (0.492)	85	0.60 (0.493)	175	0.33 (0.472)
	First-and-final Passage	89	0.74 (0.440)	75	0.67 (0.475)	86	0.52 (0.502)	182	0.38 (0.488)
	Left-and-right Passage	70	0.79 (0.413)	70	0.66 (0.478)	85	0.58 (0.497)	175	0.42 (0.495)
	Sentence-with-a-blank	90	0.71 (0.456)	76	0.57 (0.499)	86	0.60 (0.492)	182	0.45 (0.499)
Phrase Blank	*Original Passage*	89	0.57 (0.497)	76	0.34 (0.478)	85	0.26 (0.441)	182	0.24 (0.426)
	First-and-final Passage	70	0.47 (0.503)	71	0.34 (0.476)	83	0.39 (0.490)	175	0.27 (0.447)
	Left-and-right Passage	89	0.48 (0.503)	76	0.28 (0.45)	86	0.28 (0.451)	182	0.29 (0.456)
	Sentence-with-a-blank	70	0.39 (0.490)	70	0.31 (0.468)	85	0.25 (0.434)	175	0.25 (0.432)
Clause Blank	*Original Passage*	73	0.74 (0.442)	77	0.48 (0.503)	85	0.35 (0.481)	170	0.27 (0.446)
	First-and-final Passage	80	0.56 (0.499)	75	0.32 (0.47)	88	0.32 (0.468)	188	0.36 (0.480)
	Left-and-right Passage	84	0.68 (0.470)	67	0.45 (0.501)	85	0.39 (0.490)	187	0.32 (0.468)
	Sentence-with-a-blank	75	0.47 (0.502)	69	0.48 (0.503)	81	0.28 (0.454)	169	0.32 (0.468)

For the statistical analysis, the multilevel logistic regression was practiced. The multilevel models consisted of a fixed part and a random part. The fixed part considered the predictor variables which were the amount of textual information, blank type, and the English proficiency level while the random part of the model contained the nesting of variables in the data which was each individual. Table 4.2 is the summary of the fixed effects of the multilevel logistic models.

Table 4.2

Summary of Model 1 and Model 2

Predictor variables	Model 1		Model 2 (final)	
	F	Sig.	F	Sig.
Corrected Model	7.380	.000***	16.637	.000***
Amount of textual information	2.032	.107		
Blank type	98.284	.000***	104.166	.000***
English Proficiency level	39.428	.000***	35.012	.000***
English Proficiency level × Amount of textual information	2.478	.008**	2.189	.010*
English Proficiency level × Blank type	3.746	.001**	3.683	.001**
Amount of textual information × Blank type	1.422	.202		
English Proficiency level × Amount of textual information × Blank type	.981	.478		

*p<.05, **p<.01, ***p<.001

In the first model of multilevel logistic, all predictors and interactions were entered. The result of the Model 1 reports that *amount of textual information*, *amount of textual information × blank type*, and *English proficiency level × amount of textual information × blank type* were not significant predictors while *blank type*, *English proficiency level*, *English proficiency level × amount of textual information*, and *English proficiency level × blank type* were proven to be significant (p<.05). Since the goal of the logistic regression is to find the best fitting model to explain the dependent variable, the second model was formed including only significant predictors (*i.e.*, blank type, English proficiency level,

English proficiency level × amount of textual information, and English proficiency level × blank type) as Landau and Everitt (2004) advised.

Table 4.3 is the classification table which accounts for prediction of the correct answer. The classification table suggests the accuracy of the logistic model: how accurately the model predicts the dependent variable (White, 2013). The classification of the correct for incorrect answer was much higher than the classification of correct for the incorrect answers. The classification table indicates that the model classification was 83.2% correct for the incorrect answers and 47.8% correct for the correct answers. The predictors presented in the study might be more suitable for predicting the incorrect answers while more predictors are needed to be included in the model to predict the correct answers. The overall fit of the model yielded 68.4% correct classification.

Table 4.3
Prediction of the Correct Answer (Model 2)

	Predicted (N)		
Observed	Incorrect answer	Correct answer	Percentage correct
Incorrect answer	2406	485	83.2%
Correct answer	1092	1000	47.8%
Overall percentage			68.4%

The final model of multilevel logistic regression which includes only significant predictors and interactions is presented in Table 4.4. Once again, the *Original Passage*, *One-word Blank*, and the Level 4 was set as a reference category for each predictor.

Table 4.4

Final Model (Model 2) of Multilevel Logistic Regression

Variable	Level	Odds	SE	Sig.	Odds ratio
Blank type	*One-word Blank*	0			1
	Phrase Blank	-.639	.1228	.000	.528***
	Clause Blank	-.362	.1107	.001	.696**
English Proficiency level	Level 4	0			1
	Level 3	1.004	.2098	.000	2.73***
	Level 2	1.229	.2167	.000	3.417***
	Level 1	2.258	.2359	.000	9.565***
English Proficiency level by Amount of textual information	Level 4×*Original Passage*	0			1
	Level 4×*First-and-final Passage*	.284	.1384	.036	1.329*
	Level 4×*Left-and-right Passage*	.317	.1347	.020	1.372*
	Level 4×*Sentence-with-a-blank*	.294	.1312	.032	1.341*
	Level 3×*Original Passage*	0			1
	Level 3×*First-and-final Passage*	.007	.1962	.970	1.007
	Level 3×*Left-and-right Passage*	.052	.1719	.783	1.053
	Level 3×*Sentence-with-a-blank*	-.103	.1930	.587	.902
	Level 2×*Original Passage*	0			1
	Level 2×*First-and-final Passage*	-.160	.1823	.424	.852
	Level 2×*Left-and-right Passage*	-.078	.1870	.698	.925
	Level 2×*Sentence-with-a-blank*	-.112	.1795	.578	.894
	Level 1×*Original Passage*	0			1
	Level 1×*First-and-final Passage*	-.520	.222	.011	.595*
	Level 1×*Left-and-right Passage*	-.254	.1756	.211	.776
	Level 1×*Sentence-with-a-blank*	-0.830	.209	.000	.436***
English Proficiency level by Blank type	Level 4×*One-word Blank*	0			1
	Level 4×*Phrase Blank*	-.639	.1228	.000	.528***
	Level 4×*Clause Blank*	-.362	.1107	.001	.696**
	Level 3×*One-word Blank*	0			1
	Level 3×*Phrase Blank*	-.580	.1863	.004	.560**
	Level 3×*Clause Blank*	-.648	.1828	.001	.523**
	Level 2×*One-word Blank*	0			1
	Level 2×*Phrase Blank*	-.669	.2266	.002	.512**
	Level 2×*Clause Blank*	-.446	.1942	.030	.640*
	Level 1×*One-word Blank*	0			1
	Level 1×*Phrase Blank*	-.667	.2096	.002	.514**
	Level 1×*Clause Blank*	-.362	.1778	.087	.696

*p<.05, **p<.01, ***p<.001

As shown in Table 4.4, for blank type and English proficiency level variables, the probability for correct answer on the reference category was significantly different from those of other categories. Also, the probabilities for correct answer of each amount of textual information and probabilities for correct answer of each blank type were significantly different depending on the English proficiency level. Detailed analyses of the Table 4.4 are discussed in the following sections.

4.1.1 Effects of Amount of Textual Information

According to the Model 1 as in Table 4.2, the amount of textual information was not a meaningful predictor on the probability of getting correct answer. The probability for the correct answer was not significantly different depending on the amount of textual information. Therefore, the hypothesis for the first research question, 'The probability for correct answer would be significantly different depending on the amount of textual information provided in the passage' was rejected. However, the interaction effects of the amount of textual information and the English proficiency level were found, which are discussed in Section 4.1.4.

4.1.2 Effects of Blank Type

The blank type was a significant predictor variable on the probability of getting a question correct ($p<.001$). The probability for the correct answer was different depending on the blank type. Thus, the second hypothesis 'There would be some differences among the three question types in terms of probability for

correct answer' was supported by the present study. The probability for the correct answer was the highest in the *One-word Blank* questions, followed by the *Clause Blank* and the *Phrase Blank* questions. The high difficulty of Phrase Blank questions confirmed the result of Jin and Park (2004). There were also the interaction effects of blank type and English proficiency level.

4.1.3 Effects of English Proficiency Level

As illustrated in Table 4.2, the English proficiency level was a significant predictor variable on the probability of getting correct answer of the fill-in-the-blank questions. The higher the English proficiency level, the higher the probability for the correct answer was. Thus, the third hypothesis 'Differences in the probability for correct answer among the English proficiency level groups will exist even using this type of questions' was also confirmed by the present study. The English proficiency level (mock CSAT English level) that the participants achieved one month prior to the main study had a strong relationship with the probability for correct answer on the fill-in-the-blank questions in the current study.

4.1.4 Interaction Effects of Amount of Textual Information and English proficiency level

The fourth hypothesis was proven to be true by the result of statistical analysis. Refer to the hypothesis again:

> There would be interaction effects of the amount of textual information and the English proficiency level. More advanced groups will presumably prefer longer texts to shorter ones, and their correct answer rate will be higher. Less advanced groups will prefer shorter passages, and the probability for correct answer will be higher in shorter text than in longer ones.

The interaction effects of the amount of textual information and the English proficiency level was significant both in the Model 1 and the Model 2 ($p<.05$) as presented in Table 4.2. The effect of the amount of textual information was significantly different depending on the English levels. The Level 1 group with the highest English proficiency and the Level 4 with the lowest English proficiency presented the opposite trends; the lowest ability learners' probabilities for correct answer was the lowest on the longest text, on which the highest ability learners displayed the highest probabilities for correct answer. While the Level 1 participants are predicted to perform the highest on the *Original Passage*, their probabilities for the correct answer on the *Left-and-right Passage* were not statistically different from those on the *Original Passage*. Figure 4.2 depicts the interaction effects of the amount of textual information and the English proficiency level on the correct answer.

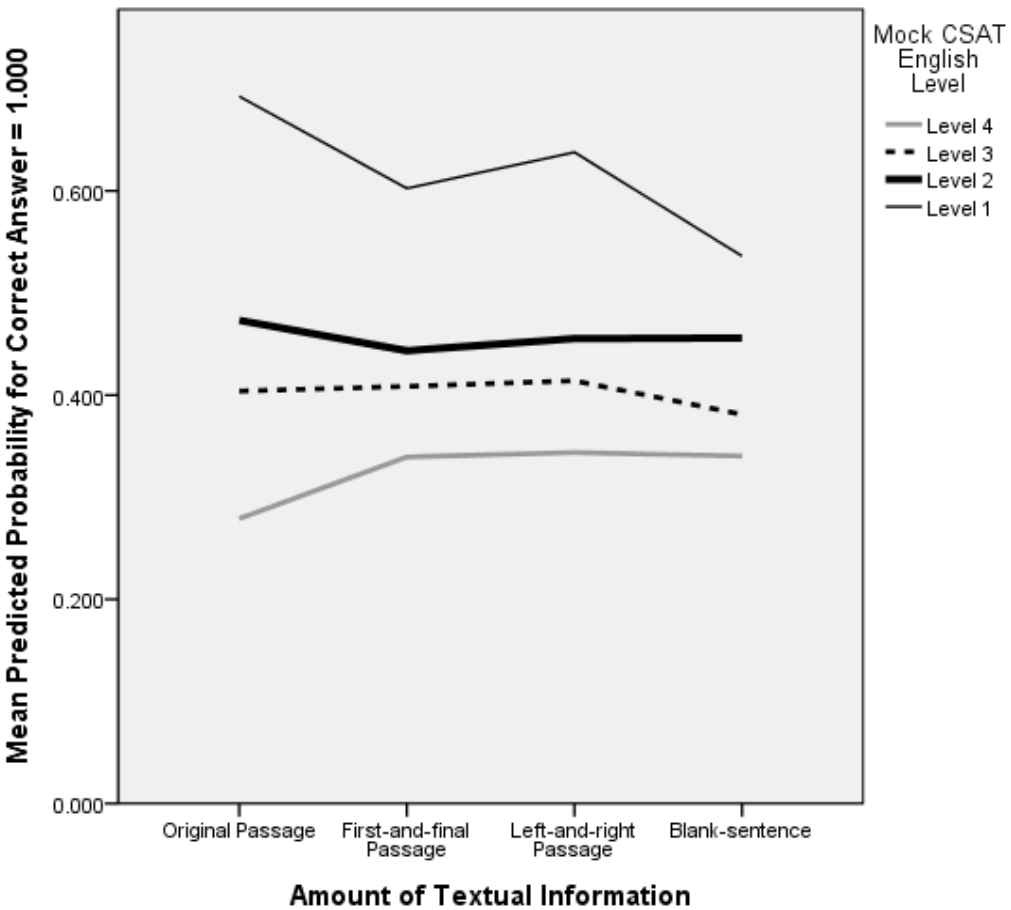

Figure 4.2 Interaction effects of amount of textual information and English proficiency level on correct answer

The Level 4 participants' probability for correct answer was significantly low on the longest *Original Passage* compared with the other text conditions. The odds of correct answer on the shorter *Sentence-with-a-blank*, *Left-and-right Passage* and the *First-and-final Passage* are increased about 130% ($p<.05$) compared with the *Original Passage* (see Table 4.4). The odds ratio of the correct answer was the highest on the *Left-and-right Passage* (odds ratio=1.372)

followed by the *Sentence-with-a-blank* (odds ratio=1.341) and the *First-and-final Passage* (odds ratio=1.329) compared with the *Original Passage*. The *Original Passage* with much information might have been too long to read and comprehend for the Level 4 learners. Since their reading fluency is not sufficient to read the whole text within the time limit, the extra context provided in the *Original Passage* might even make them walk into the trap; they might choose the wrong alternative that seemed to be the correct answer only on the word-level. In the end, reading the whole passage might not be helpful but somewhat disturbing for the Level 4 learners to reach the correct answer. The Level 2 and Level 3 participants' probabilities for correct answer were not significantly different regardless of the amount of textual information.

For the Level 2, 3 and 4 participants, the longer texts did not increase the possibility for the correct answer on the fill-in-the-blank questions. According to the result of 2018 CSAT, only 10% of the test-takers' score was 90 or higher, which is a cut-off score of Level 1 in the present study. Thus, in the real test-taking setting, reading the entire passage might not be more helpful than reading parts of the passages for about 90% of the test-taking population. The partial reading as in the *Sentence-with-a-blank*, *Left-and-right Passage*, and the *First-and-final Passage* might be even more favorable in the real test-taking context since the test-takers can save time to answer other questions.

On the contrary, the Level 1 participants showed the highest probability for correct answer on the *Original Passage* on which the Level 4 participants had the lowest probability for correct answer. On the *Sentence-with-a-blank* and the *First-and-final Passage*, the odds ratio of the correct answer is decreased by .436 ($p<.001$) and .595 ($p<.05$) respectively compared with the *Original Passage*. Unlike other groups of participants, the Level 1 learners properly use the extra context provided in the *Original Passage* to reach the answer. Only the Level 1 participants seem to be able to complete the reading task required in the reading comprehension test. However, surprisingly, their probability for correct answer on the *Left-and-right Passage* was not significantly different from that on the *Original Passage*. The three core sentences in the *Left-and-right Passage* were still useful to the Level 1 participants as well as to the learners of the other level groups. Given that the high performance can be predicted by reading only three important sentences, as well as on the full text, even the Level 1 participants who can utilize the extra context do not need to read the full text in responding to the fill-in-the-blank questions.

The different effects of amount of textual information depending on the English proficiency level may derive from the characteristics of each text condition. First, the *Sentence-with-a-blank* which has a single sentence was as useful as the *Left-and-right Passage* and the *First-and-final Passage* for the learners of the Level 2, 3, and 4 group and even more useful than the *Original*

Passage for the Level 4 learners. Refer to the *Sentence-with-a-blank* version of item No. 7.

Item No. 7 (Sentence-with-a-blank)
This is why [_____], which is why ice cream makers add stacks of sugar—as you can tell all too clearly when ice cream melts.
(a) ice cream tastes better when tea flavors are added

(b) <u>ice cream does not taste that sweet straight from the fridge</u>

(c) they serve ice cream for dessert in Chinese restaurants

(d) it is not recommended to eat ice cream while drinking hot tea

(e) ice cream tastes sweeter especially in the winter time

(The underlined words are the correct response.)

The *Sentence-with-a-blank* presented above consists of a simple cause and effect structure. Moreover, the keyword '*sugar'* and '*melts'* are closely related to '*sweet'* and '*from the fridge'* in the correct response.

Second, the *Left-and-right Passage* which has three core sentences was a helpful text condition for all level groups. Reading the *Left-and-right Passage* might suggest the high probability for correct answer on the fill-in-the-blank questions regardless of the English level. The *Left-and-right Passage* version of item No. 7 is provided below.

Item No. 7 (Left-and-right Passage)
Apparently, the higher the temperature, the more intense the flavor. This is why [_____], which is why ice cream makers add stacks of sugar—as you can tell all too clearly when ice cream melts. In a similar way, some bitter tastes, like tea, taste better when hot because they are more intense.

(a) ice cream tastes better when tea flavors are added

(b) <u>ice cream does not taste that sweet straight from the fridge</u>

(c) they serve ice cream for dessert in Chinese restaurants

(d) it is not recommended to eat ice cream while drinking hot tea

(e) ice cream tastes sweeter especially in the winter time

In the *Left-and-right Passage*, the sentence preceding the *Sentence-with-a-blank* explains the underlying principle of the phenomenon, and the sentence following the *Sentence-with-a-blank* gives an example.

Since the *Sentence-with-a-blank* and the *Left-and-right Passage* include the core information of the text, the correct answer can be reached in the fill-in-the-blank question when reading those short texts as well as when reading the full text. When the comprehension of the entire text is expected in the question type, the required reading task might not have completed. In this sense, answering the reading comprehension questions without reading the entire text can be interpreted as a test-wiseness; the result of the test might not reflect the learners' reading ability that test-developers intend to measure. If the response is the result of the partial reading, whether the test score truly reflects the learners' reading ability is questionable even though the response itself is correct.

The problem is that the partial reading is a typical behavior of Korean high school students in dealing with the fill-in-the-blank questions as proven by the previous studies (*e.g.*, Haam, 2006; Oh, 1999, Kim, 2015). The partial reading is

known to be somewhat helpful for the better score on the fill-in-the-blank questions. The students are even instructed to do so in public schools and private English institutes. The current study proved that the partial reading in the fill-in-the-blank questions is indeed effective as the pervasive belief. The fill-in-the-blank question as a type of reading comprehension questions should be reconsidered if the test score of the question type is a product of partial reading; the interpretation of the fill-in-the-blank question score cannot reflect the participants' genuine linguistic ability as intended.

4.1.5 Interaction Effects of Blank Type and English proficiency level

The interaction of the blank type and the English proficiency level was significant both in the Model 1 and the Model 2 ($p<.05$). For all participant groups, the probability for the correct answer was the highest on the *One-word Blank* questions, followed by the *Clause Blank* and *Phrase Blank* questions. However, the sensitivity to the blank type was dissimilar depending on the differing level group. The lower the English level, the more sensitive to the effects of blank type. Thus, the fourth hypothesis 'There would be interaction effects of the blank type and the English proficiency level groups. Less advanced groups will be more sensitive to the blank type while more advanced will show similar probability for correct answer across the blank type.' was also confirmed by the present study.

The Level 4 participants were more sensitive to blank type than other groups of participants. Their probabilities for correct answer on the *Phrase Blank* and *Clause Blank* questions is significantly different from that on the *One-word Blank* questions. On the *Clause Blank* questions, the odds ratio of the correct answer was decreased by .696 ($p<.05$) compared with the *One-word Blank* questions. They exhibited even lower probability for correct answer on the *Phrase Blank* questions with an odds ratio of .528 ($p<0.01$). The effects of blank type were the strongest for the Level 4 learners; they experienced the most stratified difficulty depending on the type of words fit in the blank.

The Level 2 and Level 3 participants showed a similar trend as the Level 4 participants. For the participants of the Level 3 group, their probability for correct answer on the *Clause Blank* and the *Phrase Blank* questions was significantly lower with an odds ratio of .523 and .560 ($p<.05$) respectively in comparison with the *One-word Blank* questions. Likewise, the Level 2 participants' probability for correct answer on the *Clause Blank* and the *Phrase Blank* questions was lower with an odds ratio of .640 and .512 ($p<.05$) respectively compared with the *One-word Blank* questions. Since the odds ratio for correct answer on the *Clause Blank* and the *Phrase Blank* questions were close, the logistic model which reset the *Phrase Blank* as a reference category was presented for detailed investigation (see Table 4.5). When the *Phrase Blank* question was set as a baseline, there was no significant difference between the

Level 2 and Level 3 participants' probability for correct answer on the *Clause Blank* and the *Phrase Blank* questions.

Table 4.5

Interaction Effects of Blank Type and English proficiency level

Variable	Level	Odds	SE	Sig.	Odds ratio
One-word Blank as a reference category	Level 4×*One-word Blank*	0			1
	Level 4×*Phrase Blank*	-.639	.1228	.000	.528***
	Level 4×*Clause Blank*	-.362	.1107	.001	.696**
	Level 3×*One-word Blank*	0			1
	Level 3×*Phrase Blank*	-.580	.1863	.004	.560**
	Level 3×*Clause Blank*	-.648	.1828	.001	.523**
	Level 2×*One-word Blank*	0			1
	Level 2×*Phrase Blank*	-.669	.2266	.002	.512**
	Level 2×*Clause Blank*	-.446	.1942	.030	.640*
	Level 1×*One-word Blank*	0			1
	Level 1×*Phrase Blank*	-.667	.2096	.002	.514**
	Level 1×*Clause Blank*	-.362	.1778	.087	.696
Phrase Blank as a reference category	Level 4×*Phrase Blank*	0			1
	Level 4×*One-word Blank*	0.639	0.1163	.000	1.894***
	Level 4×*Clause Blank*	0.276	0.1189	.020	1.318**
	Level 3×*Phrase Blank*	0			1
	Level 3×*One-word Blank*	0.58	0.201	.004	1.787**
	Level 3×*Clause Blank*	-0.067	0.2054	.743	0.935
	Level 2×*Phrase Blank*	0			1
	Level 2×*One-word Blank*	0.669	0.2118	.002	1.953**
	Level 2×*Clause Blank*	0.224	0.2121	.292	1.251
	Level 1×*Phrase Blank*	0			1
	Level 1×*One-word Blank*	0.667	0.2127	.002	1.947**
	Level 1×*Clause Blank*	0.304	0.2046	.137	1.355

*p<.05, **p<.01, ***p<.001

The Level 1 participants displayed the weakest sensitivity to the blank type. Their pattern of probability for correct answer was not very different from the other level groups; the odds ratio for correct answer was the highest on the *One-word Blank* questions, followed by the *Clause Blank* and the *Phrase Blank* questions. The odds ratio of the correct answer on the *Phrase Blank* questions was decreased by .514 ($p<.05$) compared with the *One-word Blank* questions. However, the odds ratio of the correct answer on the *Clause Blank* questions compared with the *One-word Blank* questions was also not significantly different ($p>.05$). Furthermore, the odds ratio of the correct answer on the *Clause Blank* questions compared with the *Phrase Blank* questions was insignificant ($p>.05$). To sum up, the probability for correct answer difference between the blank types was a lot smaller for the Level 1 group.

According to the statistical analysis, the effects of the blank type were the strongest for the lowest ability participants. The Level 4 group's probability for correct answer showed a significant difference between the three blank types. The Level 2 and Level 3 participants are predicted to perform significantly higher on the *One-word Blank* questions, however, their probability for correct answer on the *Clause Blank* questions and the *Phrase Blank* question was not significantly different. For the Level 1 group, their probability for correct answer on the *Clause Blank* questions was not significantly different from that on the *Phrase Blank* questions. Also, the probability for correct answer difference

between the *One-word Blank* questions and the *Clause Blank* questions revealed insignificance. Thus, the higher the English ability, the weaker the effects of the blank type were.

Still, the probabilities for correct answer on the *One-word Blank* questions are much higher than other two blank types regardless of the English proficiency level. While the probabilities for correct answer depending on the blank type were not significantly different for high proficiency learners, their probability for correct answer on the *One-word Blank* question was significantly higher than those on the *Phrase Blank* question and *Clause Blank* question. The *One-word Blank* questions are composed of a single short blank and includes little more context than other questions. The following is an example of the *One-word Blank* question:

> **Item No. 1 (One-word Blank)**
> Magicians are honest deceivers. To investigate the secret used by magicians to fool their audiences, Jastrow worked with two great illusionists. He invited these performers to his laboratory and had them participate in a range of tests measuring their speed of movement and accuracy of finger motion. But Jastrow's results revealed little out of the ordinary. He demonstrated magic has little to do with fast movements. **Instead, magicians use a range of [_____] weapons to fool their audiences.** The technique of suggestion, which captures people's minds, plays a key role in the process. In the same way that people can be made to believe that they once went on a non-existent trip in a hot-air balloon, so magicians have to be able to manipulate people's perception of performance.
> (a) ethical (b) political (c) physical
> (d) economic (e) <u>psychological</u>
>
> (The underlined words are the correct response.)

Over 70% of the participants answered correctly on the example above. The amount of information in the *Sentence-with-a-blank* which is bolded in the example might be sufficient to reach the correct answer.

On the other hand, the *Phrase Blank* questions and *Clause Blank* questions were a lot more challenging than the *One-word Blank* questions irrespective of the English level. Interestingly, the *Phrase Blank* questions were more demanding for the low proficiency learners than the *Clause Blank* questions in spite of the fewer number of words required in the correct response. The examples of the *Clause Blank* and the *Phrase Blank* questions are presented below:

Item No. 2 (Clause Blank)
Kate and her classmate, Jamie, were the youngest of the winning teams in the Stop Racism National Video Competition in 1998. Their submission depicted children playing at a nursery accompanied by messages such as "Everyone is the same; all these babies are beautiful." Kate says the point of their video was to show that human beings are not genetically coded with racist attitudes. "[_____]," she explains. And she adds, "The children at the nursery don't say, 'I am not playing with you since you're black,' because they don't really know about racism." According to Kate, education begins with the younger generation. And if they grow up experiencing all sorts of cultures, they have less bias.

(a) Cultures are very important for winning the competition

(b) Different skin colors give a bad impression to babies

(c) Babies should be taken care of in domestic surroundings

(d) <u>Children pick up racist ideas from their surroundings as they grow up</u>

(e) Educational institutes are free from racism due to the government's efforts

Item No. 10 (Phrase Blank)
Journeys are the midwives of thought. Few places are more conducive to internal conversations than a moving plane, ship, or train. There is an almost peculiar correlation between what is in front of our eyes and the thoughts we are able to have in our heads: large thoughts at times requiring large views, new thoughts new places. Introspective reflections which are liable to stall are helped along by the flow of the landscape. The mind [_____] when thinking is all it is supposed to do. The task can be as paralyzing as having to tell a joke or mimic an accent on demand. Thinking improves when parts of the mind are given other tasks, are charged with listening to music or following a line of trees.

(a) <u>may be reluctant to think properly</u>

(b) may focus better on future thoughts

(c) can become confused by multitasking

(d) is likely to be paralyzed by fear of new tasks

(e) can be distracted from what is before the eyes

While 65% of the participants answered correctly on item No. 2, only 17% of the participants answered correctly on item No. 10. The *Clause Blank* question might be relatively easy because the answer is more independent from the context than the *Phrase Blank* question. It might be possible to reach the correct response just by eliminating the unattractive choices on item No. 2. By eliminating the unattractive choices, the learners do not need to comprehend the full text. For example, the distractors like '*Cultures are very important for winning the competition, Different skin colors give a bad impression to babies, Babies should be taken care of in domestic surroundings'* seem somewhat not educational. The other distractor '*Educational institutes are free from racism due to the government's efforts'* might not reflect the reality which can be interpreted as an

illogical inference. Since only the correct answer is left after eliminating the unattractive choices, the learners could get to the correct answer without even reading the text.

However, the *Phrase Blank* questions may be more dependent on the context. Furthermore, the correct answer can be only inferred from the overall understanding of the text rather than the immediate words. The result confirmed the previous study on item difficulty in CSAT. In Jin and Park (2004), *inferring the parenthesis (phrase)* was revealed as a significant predictor variable on item difficulty in CSAT while *inferring the parenthesis (words)* and *inferring the parenthesis (sentence)* were insignificant. The sensitivity to cohesion could be blamed for the high difficulty of the *Phrase Blank* questions. Since the missing information in the *Phrase Blank* is more dependent on the context, the *Phrase Blank* questions require full comprehension of the text. Thus, the Phrase Blank questions might be challenging for most of the test-takers, especially the lowest English ability learners.

4.2 Questionnaire

The questionnaire was implemented to investigate the learners' test-taking behavior. Whether the partial reading suggested as a test-wiseness strategy is truly prevalent among test-taking population was verified. Also, if there is any difference in test-taking behavior among English proficiency groups was explored. One hundred sixty-seven participants completed the questionnaire.

According to the questionnaire, although reading the text from the beginning to the end could be expected in reading comprehension test, many participants are involved in the unexpected partial reading. Table 4.6 illustrates the participants' response to Question 1, 2, 3, and 4. Since the four questions could be answered with a five-point Likert scale, one-way ANOVA was conducted to investigate if there is any relationship between the test-taking behavior and the English proficiency level. The high score was given to the sequential reading, while the low score was given to the partial or mixed order reading. Descriptive statistics of the one-way ANOVA are presented in Table 4.7.

Table 4.6
Participants' Response to Question 1, 2, 3, and 4

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Sum
Q1 I read the whole passage in each item when I take an English reading comprehension section of CSAT.	N	11	46	69	104	47	277
	%	3.97	16.61	24.91	37.55	16.97	100
Q2 I read the whole passage when I answer the fill-in-the-blank questions in CSAT.	N	18	52	71	85	51	277
	%	6.5	18.77	25.63	30.69	18.41	100
Q3 I read the questions and alternative before I read the passage when I take an English reading comprehension section of CSAT.	N	25	56	65	94	37	277
	%	13.36	33.94	23.47	20.22	9.03	100
Q4 I read the questions and alternative before I read the passage when I answer the fill-in-the-blank questions in CSAT.	N	19	51	76	84	45	277
	%	16.36	30.55	27.64	18.55	6.91	100

Table 4.7
Descriptive Statistics of One-way ANOVA

		N	Mean	Std. Deviation	S.E.	95% Confidence Interval	
						Lower	Upper
Q1	Level 4	116	2.8966	1.09051	0.10125	2.696	3.0971
	Level 3	57	2.3333	0.95119	0.12599	2.0809	2.5857
	Level 2	49	2.3673	1.09343	0.1562	2.0533	2.6814
	Level 1	55	2.1091	0.93636	0.12626	1.856	2.3622
	Total	277	2.5307	1.07843	0.0648	2.4031	2.6582
Q2	Level 4	116	3.1466	1.09746	0.1019	2.9447	3.3484
	Level 3	57	2.4386	1.14981	0.1523	2.1335	2.7437
	Level 2	49	2.3878	1.07657	0.1538	2.0785	2.697
	Level 1	55	2.0182	0.99053	0.13356	1.7504	2.286
	Total	277	2.6426	1.16969	0.07028	2.5042	2.781
Q3	Level 4	116	3.0517	1.1409	0.10593	2.8419	3.2616
	Level 3	57	2.807	1.18681	0.1572	2.4921	3.1219
	Level 2	49	2.5714	1.20761	0.17252	2.2246	2.9183
	Level 1	55	2.3455	1.09237	0.14729	2.0501	2.6408
	Total	277	2.7762	1.17972	0.07088	2.6366	2.9157
Q4	Level 4	115	2.8609	1.03353	0.09638	2.6699	3.0518
	Level 3	57	2.7544	1.19942	0.15887	2.4361	3.0726
	Level 2	48	2.5833	1.25195	0.1807	2.2198	2.9469
	Level 1	55	2.3636	1.20744	0.16281	2.0372	2.6901
	Total	275	2.6909	1.15374	0.06957	2.5539	2.8279

As presented in Table 4.8, the test-taking behavior was significantly different depending on the English proficiency level. *'Reading the whole passage in English section of CSAT'* ($p<.001$), *'Reading the whole passage dealing with the fill-in-the-blank questions'* ($p<.001$), and *'Reading the questions and alternatives before reading the passage in English section of CSAT'* ($p<.05$) *'Reading the questions and alternatives before reading the passage dealing with the fill-in-the-blank questions'* ($p=.056$) was significantly different depending on the English proficiency level.

Table 4.8
Test-taking Behavior Depending on English Proficiency Level

		SS	df	Mean Square	F	Sig.
Q1	Between Groups	28.831	3	9.610	8.980	.000***
	Within Groups	292.158	273	1.070		
	Total	320.989	276			
Q2	Between Groups	56.459	3	18.820	15.998	.000***
	Within Groups	321.158	273	1.176		
	Total	377.617	276			
Q3	Between Groups	21.120	3	7.040	5.294	.001**
	Within Groups	363.003	273	1.330		
	Total	384.123	276			
Q4	Between Groups	9.998	3	3.333	2.546	.056
	Within Groups	354.729	271	1.309		
	Total	364.727	274			

*p<.05, **p<.01, ***p<.001

In addition, the test-taking behavior of the Level 4 participants is significantly different from other level groups, according to post-hoc results of multiple comparisons, using the Bonferroni procedure (see Appendix B). More Level 1, 2, and 3 participants admitted that they do not read the entire passage when they answer the general English section of CSAT (p<.05) and the fill-in-the-blank questions (p<.001) than the Level 4 participants. More Level 1 participants said that they read the questions and alternative before they read the text than the Level 4 participants (p<.05). In other words, the Level 4 participants with low linguistic ability are less accustomed to test-wiseness strategies of partial reading while other participant groups, especially Level 1

participants, are more familiar with the strategic approach to the general CSAT and the fill-in-the-blank questions.

The result accords with the previous studies on test-wiseness strategies. In second language research, high proficiency learners were observed to use more various test-wiseness strategies than learners with low proficiency (Ghafournia & Afghari, 2013; Haam, 2006; Nam, 2015). However, just utilizing many strategies does not automatically guide to correct response; the strategies that are adequate for the test format and test setting rather than just a variety of strategies influence on the better performance of test-takers (Dolly & William, 1986; Ghafournia & Afghari, 2013; Nevo, 1989). The high ability learners in the present study reported to use the effective test-wiseness strategies for the question type (*i.e.*, partial reading and mixed order reading) which were the main concern of the present study more frequently than the low ability learners.

According to the main study, the learners with low proficiency were not predicted to perform better by reading the full texts. Rather, most of the learners might benefit from the reading a few core sentences as a test-wiseness strategy. While the low proficiency learners' reading fluency is expected to be lower, they seem to spend more time reading sentences that are not essential. Both the lack of the linguistic ability and test-wiseness strategies might cause a low performance on fill-in-the-blank questions.

Also, the participants who tend to read the whole passage in the fill-in-the-blank questions got a low score. The correlation between the score in Question 2 and the total score of the main task was calculated. The score of the fill-in-the-blank questions and the normal reading behavior reflected by the questionnaire showed the negative correlation ($r = -.23$, $p < .001$). In other words, the more participants read the passage, the lower their scores were. Reading the whole passage influenced their score negatively while the test developers might expect the participants to read the entire passage. The result directly connects to the heart of the main study; reading the entire passage might only result in lower performance in dealing with the fill-in-the-blank questions. The fill-the-blank question is an awkward kind of reading comprehension question which the test-takers should not read the full text to get a high score.

As shown in Table 4.9, over a half of the participants admitted they do not read from the first sentence of the passage when they answer the fill-in-the-blank questions. It is confirmed that the majority of participants are involved in partial or mixed order reading behavior. About fifty percent of the participants reported that they read the *Sentence-with-a-blank* or the *Left-and-right Passage* first. The reasons for this reading behavior were mainly the time pressure, item difficulty, and instruction. As discussed earlier, the passages presented in the fill-in-the-blank questions are too long and difficult, considering reading fluency of Korean EFL high school students. Since the CSAT is the most important test for

the students, the extreme difficulty results in the unexpected reading process. Due to the strong washback effects of the CSAT, even teachers might have adopted test-wiseness strategies in their instruction. As shown in Table 4.9, many of the participants said they were taught to use test-wiseness strategies in dealing with fill-in-the-blank question. The instructed test-wiseness strategies might reproduce the overuse of test-wiseness strategies instead of improving the learners' reading ability.

Table 4.9
Participants' Response to Question 5 and 6

Q5 I read [ ] first when I answer the fill-in-the-blank questions in CSAT.	N	%
The first sentence of the passage	126	45.81
The final sentence of the passage	9	3.27
The left sentence of the blank-sentence	58	21.09
The blank-sentence	61	22.18
The right sentence of the blank sentence	18	6.05
Total	275	100
Q6 I don't read the passage in the fill-in-the-blank questions from the beginning because of [ ].	N	%
The time pressure	67	40.11
The item difficulty	51	30.5
The instruction	49	29.3
Total	167	100

Chapter 5. Conclusion

This chapter summarizes the major findings of the current study. The implication of the present study is also described, followed by the limitation and suggestion for further research.

5.1 Summary of Major Findings

The present study investigated Korean EFL high school learners' performance on the fill-in-the-blank questions in Korean CSAT, focusing on the amount of textual information, blank type, and the English proficiency level. 279 Korean high school students participated in the main task and answered the questionnaire to delve into the predictor variables on the probability of getting correct answer of the fill-in-the-blank questions. The current study answered the research questions presented in the introduction chapter. The first three research questions aim to investigate the significant predictor variables on the probability of getting a question correct. The blank type and the English proficiency level were significant predictors while the amount of textual information was not a significant predictor. The other two research questions are concerned with the interaction effects which lead to the two major findings.

First, reading the three core sentences might draw the highest performance of the learners while reading the entire text may not be always helpful for all levels of learners. Contrary to the hypothesis, the full text is not useful for the learners with the lowest English ability since their reading fluency

is not sufficient to read and understand the whole passage within the time limit. For the two mid-level groups in the present study (*i.e.*, Level 2 and Level 3), their probability for correct answer would remain similar irrespective of the given amount of textual information. The learners with the highest English ability are the only ones who can utilize the extended context to reach the correct answer; their probability for correct answer was the highest when the full text was provided. However, even for the highest ability group, their probability for correct answer on the three core sentences (*i.e.*, *Left-and-right Passage*) was not significantly different from the probability for correct answer on the full text. Reading the entire text from the beginning to the end does not guarantee the highest probability for correct answer for all level learners. Rather, reading only three sentences which have the most important information might be beneficial for all groups of learners. As long as reading the three sentences is more helpful than (or as helpful as) reading the entire text, it is meaningless to read the whole passage to answer the fill-in-the-blank questions.

Second, the learners with lower ability might be more sensitive to the blank type while the learners with higher ability are less sensitive to the effects of the blank type. The blank type was a significant predictor on the probability of getting a question correct, for all level groups; the *Phrase Blank* questions were the most demanding ones, followed by the *Clause Blank* and the *One-word Blank* questions. However, the relative sensitivity to the blank type increased as

the English level decreased. The lowest level group showed a significant difference in probability for correct answer between the blank types. The two mid-level group's probability for correct answer on the *Phrase Blank* questions was not significantly different from the probability for correct answer on the *Clause Blank* questions while those probabilities for correct answer were significantly different from the probability for correct answer on the *One-word Blank* questions. The highest level group's sensitivity to the blank type was the weakest; their probability for correct answer on the *Clause Blank* questions was not significantly different from either on the *Phrase Blank* questions or on the *One-word Blank* questions. The blank type gap for the highest ability learners was only found between the *One-word Blank* questions and *Phrase Blank* questions. The *Clause Blank* questions are more demanding than the *One-word Blank* questions because they require more reading. The *Phrase Blank* questions are more challenging than the *Clause Blank* questions since the missing information is more dependent on the context. The low ability learners easily fail to read longer words required in the blank as in the *Clause Blank* questions and *Phrase Blank* questions. Also, they are less likely to fully comprehend the text to find the missing information which are dependent on the context whereas the high ability learners manage to do so.

5.2 Implication of the Findings

The findings of the current study might question what fill-in-the-blank questions try to measure as a reading comprehension question since the question type is susceptible to test-wiseness. The fill-in-the-blank question's vulnerability to test-wiseness also draws the washback effects of the fill-in-the-blank questions in the high-stakes CSAT. Two pedagogical implication are suggested to the test developers and teachers.

First, the test developers are encouraged to keep the goal of the fill-in-the-blank questions in mind. The result of the questionnaire form the study implies that the learners mostly do not read the text in the fill-in-the-blank questions. Over half of the participants said they do not read the whole passage when they deal with the test item. It is problematic when the learners do not read in a reading comprehension test. The problem gets worse when they can still get the same score without reading the text. As long as learners can get away with the partial reading, the fill-in-the-blank question cannot properly function as a measure of reading comprehension.

Second, teachers are advised to spare more time on fostering reading comprehension ability instead of promoting test-wiseness strategies. In reality, many students waste their time practicing the test-wiseness strategies just to get a high score. The school also reproduce the partial reading or mixed order reading; some teachers teach test-wiseness strategies, depriving students of opportunity to

foster genuine linguistic skill. The heart of English education in school is to enhance communicative competence, not to get a high score in the CSAT. When teachers teach test-wiseness strategies, not English, it is natural that many Korean EFL learners fail to acquire genuine communicative competence in spite of investment of a lot of time, money, and effort (Choi, 2008). Admitting the strong washback effects of the CSAT, teachers are recommended to spend more time teaching English rather than test-wiseness strategies.

5.3 Limitation and Suggestion for Future Research

Despite the significance and implications mentioned in the previous section, the current study might bear some limitation. The cognitive process of taking the fill-in-the-blank questions cannot be assured by the amount of textual information presented in the current study. Even though the *Original Passage* was given as a control condition, some participants might have still used the test-wiseness strategies; they might have read only parts rather than the whole passage even in the *Original Passage*. In an attempt to mitigate the limitation of the present study, other text conditions which would reflect test-wiseness strategies can be suggested and compared with the result of the study. Other online research methods like eye-tracking can be adopted to scrutinize participants' experience in dealing with the fill-in-the-blank questions.

Finally, if the performance on the fill-in-the-blank questions differs depending on English learning experience and native language can be the

interest of further research. The partial reading is a test-taking behavior also found in L1 and ESL research. However, other population groups might perform differently on the fill-in-the-blank questions because it is a very much Koreanized form of fill-in-the-blank question. Allan (1992) offered that the use of test-taking strategy might differ among the student population group. Further research might explain if the success with the little amount of textual information is more relevant to Korean EFL learners than other population groups.

References

Akbari, R. (2012). Validity in language testing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 30-36). New York: Cambridge University Press.

Alderson, J. C. (1979). The cloze procedure and the mock CSAT English level in foreign language. *TESOL Quarterly, 13*(2), 219-223.

Alderson, J. C. (2000). *Assessing reading.* Cambridge University Press.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridege: Cambridge University Press.

Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing, 9*(2), 101-119.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, DC: National Academy of Education.

Andreassen, R., & Bråten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading, 33*(3), 263-283.

Beach, K. L. (2008). *The effect of media, text length, and reading rates on college student reading comprehension levels*. Unpublished Doctoral dissertation. University of Phoenix, U.S.A.

Bachman, L. (1982). The trait structure of cloze test scores. *TESOL Quarterly, 16*(1), 61-70.

Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly, 19*(3), 535-556.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Bensoussan, M., & Ramraz, R. (1984). Testing EFL reading comprehension using a multiple-choice rational cloze. *The Modern Language Journal, 68*(3), 230-239.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, *64*(3), 311-317.

Bussis, A., & Chittenden, E. (1987). Research currents: What the reading tests neglect. *Language Arts, 64*(3), 302-308.

Carter, M. D., Walker, M. M., O'Brien, K., & Hough, M. S. (2017). The effects of text length on reading abilities in accelerated reading tasks. *Speech, Language and Hearing,* DOI: 10.1080/2050571X.2017.1415086. 1-11.

Cha, K. A. (1995). The effect of text length in the diagnosis of reading comprehension. *Korean Journal of Applied Linguistics*, 8. 249-275.

Chapelle, C. A., & Abraham, R. G. (1990). Cloze method: What difference does it make? *Language Testing*, *7*(2), 121-146.

Chang, K. (2004). A model of predicting item difficulty of the reading test of College Scholastic Ability Test. *Foreign Languages Education, 11*(1), 111-130.

Cha´vez-Oller, M.A., Chihara, T., Weaver, K.A., & Oller, J.W. Jr. (1994). When are cloze items sensitive to constraints across sentences? In Oller, J.W. Jr., & Jonz, J., (Eds), *Cloze and coherence*. London: Associated University Press, 229-245.

Chihara, T., Oller J., Weaver, K., & Chavez-Oller, M. (1977). Are cloze items sensitive to constraints across sentences? *Language Learning, 27*(1), 63-73.

Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25*(1), 39-62.

Chon, Y. V., & Shin, T. (2010). Item difficulty predictors of a multiple-choice reading test. *English Teaching, 65*(4), 257-282.

Chujo, K., & Utiyama, M. (2005). Understanding the Role of Text Length, Sample Size and Vocabulary Size in Determining Text Coverage. *Reading in a Foreign Language, 17*(1), 1-22.

Cohen, A. D. (1984). On taking language tests: What the students report. *Language Testing, 1*(1), 70-81.

Cohen, A. D. (2012). Test-taking strategies. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 96-104). New York: Cambridge University Press.

Commander, N. E., & Stanwyck, D. J. (1997). Illusion of knowing in adult readers: Effects of reading skill and passage length. *Contemporary Educational Psychology, 22*(1), 39-52.

Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General, 130*(2), 208.

Dolly, J. P., & Williams, K. S. (1986). Using test-taking strategies to maximize multiple-choice test scores. *Educational and Psychological Measurement, 46*(3), 619-625.

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *arXiv preprint*. Retrieved from arXiv:1701.04858.

Farr, R., Pritchard, & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice comprehension test. *Journal of Educational Measurement, 27*(3), 209-226.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching. *American Psychologist, 39*, 193-202.

Ghafournia, N., & Afghari, A. (2013). The interaction between cognitive test-taking strategies, reading ability, and reading comprehension test performance of Iranian EFL Learners. *English Language Teaching, 6*(8), 139-150.

Grabe, W. (2009). *Reading in a second language*. New York, NY: Cambridge University Press.

Greenhouse, J. B., Bromberg, J. A., & Fromm, D. (1995). An introduction to

logistic regression with an application to the analysis of language recovery following a stroke. *Journal of communication disorders, 28*(3), 229-246.
Haam, E. (2006). The correlation between test-taking strategy use and reading comprehension test. *English Teaching, 61*(3), 3-28.
Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M. M., Butler, F. A., & Oller Jr, J. W. (1988). Multiple-choice cloze items and the test of English as a foreign language. *ETS Research Report Series, 1988*(1), i-117.
Han, M. (2007). Test-taking strategies for the cloze by skilled and less skilled EFL readers. *The New Korean Journal of English Language and Literature, 49*(1), 263-283.
Heck, R. H., Tabata, L., & Thomas, S. L. (2013). *Multilevel and Longitudinal Modeling with IBM SPSS*. Abingdon: Routledge.
Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: A Wiley-Interscience Publication.
Hughes, A. (1989). *Testing for language teachers*. New York: Cambridge University Press.
Jin, K. A., & Park, C. (2004). The prediction of English item difficulty in college scholastic ability test. *English Teaching, 59*(1), 267-278.
Jin, K. A., & Park, C. (2005). Factors related to the difficulty of College Scholastic Ability Test English items. *English Teaching, 60*(3), 17-25.
Johnston, P. (1987). Assessing the product, and the process of assessment in the language arts. In J. Squire (Ed.), *The dynamics of language learning*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.
Jonz, J. (1990). Another turn in the conversation: What does cloze measure? *TESOL Quarterly, 24*(1), 61-83.
Katz, S., Lautenschlager, G. J., Blackburn, A. B., & Harris, F. H. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science, 1*(2), 122-127.
Kibby, M. W. (1980). Intersentential processes in reading comprehension. *Journal of Reading Behavior, 12*(4), 299-312.
Kim, M. J. (2001). Comparing question types in the high school English textbooks and the College Scholastic Abilities Test. *English Teaching, 56*(4), 349-371.
Kim, K. (2014). *Jujemumui balhyeonyangsange ttara binkanui naeyongeul churonhaneun bangan* (*A study of the methods of inferring the words missing in the blank from the ways topic sentences work in the paragraphs*). Unpublished MA thesis, Inje University, Gimhae.
Kim, Y. (2015). The retrospective and perceptive: The 20 years of College Scholastic Ability Test (CSAT). *Studies in British and American Language and Literature, 119*, 275-199.
Kim, Y., & Ma, J. H. (2012). Text readability of the college scholastic ability test and high school English II textbooks. *English Teaching, 67*(4), 195-218.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A

construction-integration model. *Psychological Review, 95*(2), 163.
Kintsch, E., & Kintsch, W. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 71-92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review, 85*(5), 363.
Kwon, O. (2015). A history of policies regarding the English section of Korea's College Scholastic Ability Test. *English Teaching, 70*(5), 3-34.
Landau, S., & Everitt, B. S. (2004). *A Handbook of statistical analyses using SPSS*. NY: CHAPMAN & HALL/CRC.
LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815-852.
Lee, J. W. (2002). An analysis of test-taking strategies for the cloze. *English Teaching, 57*(1), 213-237.
Markham, P. L. (1985). The rational deletion cloze and global comprehension in German. *Language learning, 35*(3), 423-430.
Mayher, J., & Brause, R. (1986). Learning through teaching: Is testing crippling integrated language education? *Language Arts, 63*(4), 390-3.
McKenna, M. C., & Layton, K. (1990). Concurrent validity of cloze as a measure of intersentential comprehension. *Journal of Educational Psychology, 82*(2), 372.
Mehrpour, S., & Riazi, A. (2004). The impact of text length on EFL students' reading comprehension. *Asian EFL Journal, 6*(3), 1-13.
Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher, 18*(2), 5-11.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist, 50*(9), 741.
Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45*(1), 35-44.
Nam, Y. K. (2015). Korean EFL students' strategy use in fill-in-the-blank questions. *English Teaching, 70*(4), 23-43.
Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing, 6*(2), 199-215.
Nosrati, V. (2015). Reading test-taking strategies in general training IELTS. *Advances in Language and Literary Studies, 6*(5), 134-142.
Oh, J. (1999). Blank-filling and reading strategies. *English Teaching, 54*(2), 85-108.
Ohnmacht, F. W., Weaver W. W., & E. T. Kohler. (1970). Cloze and closure: A factorial study. *The Journal of Psychology*, *74*, 205-217.
Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the

difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods, 40*(4), 1001-1015.

Palmer, R. F., Graham, J. W., White, E. L., & Hansen, W. B. (1998). Applying multilevel analytic strategies in adolescent substance use prevention research. *Preventive Medicine*, *27*(3), 328-336.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading, 11*(4), 357-383.

Perfetti, C. A., & Britt, M. A. (1995). Where do propositions come from. In C. A. Weaver. III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch*, 11-34. Abingdon: Routledge.

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology, 97*(1), 70.

Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension. *Current Directions in Psychological Science, 14*(5), 276-279

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. J. (2012). Psychology of reading. New York, NY: Psychology Press.

Rothkopf, E. Z. (1965). Some theoretical and experimental approaches to problems in written instruction. *Learning and the Educational Process*, 965.

Rothkopf, E. Z., & Billington, M. J. (1983). Passage length and recall with test size held constant: Effects of modality, pacing, and learning set. *Journal of Memory and Language. 22*(6). 667.

Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23*(4), 441-474.

Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing, 17*(1), 85-114.

Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*. *17*(2). 229-255.

Shin, S. (2018). The characteristics of absolute grading of the College Scholastic Ability Test English Section. *English Teaching, 73*(3), 95–113.

Smith, J. K. (1982). Converging on correct answers: A peculiarity of multiple choice items. *Journal of Educational Measurement*, *19*(3), 211-220.

Song, H. (1998). The effect of the College Scholastic Ability Test (CSAT) on Korean college students' reading processes and their perceptions of reading in English. *English Teaching, 53*(4), 265-290.

Soudek, M., & Soudek, L. I. (1983). Cloze after thirty years: new uses in language teaching. *ELT Journal, 37*(4), 335-340.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, *30*(4), 415-433.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Weaver, W. W., & A. J. Kingston. (1963). A factor analysis of cloze procedure and other measures of reading and language ability. *The Journal of Communication, 13*(4), 252-261.

White, J. L. (2013). Logistic regression model effectiveness: Proportional chance criteria and proportional reduction in error. *Journal of Contemporary Water Research & Education*, *2*(1), 4-10.

Yamashita, J. (2003). Processes of taking a blank-filling test: comparison of skilled and less skilled EFL readers. *Language testing, 20*(3), 267-293.

Yi, Y. S. (2013). On the optimal text length of reading comprehension tests. *Journal of The English Language and Literature, 55*(4), 505-530.

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current directions in psychological science*, *8*(1), 15-18.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin, 123*(2), 162.

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. J. Traxler, & M. A. Gernsbacher (Eds). *Handbook of Psycholinguistics*, (pp. 725-764). Amsterdam: Elsevier.

Appendices

Appendix A: Material

The Original Test Items

1. Magicians are honest deceivers. To investigate the secret used by magicians to fool their audiences, Jastrow worked with two great illusionists. He invited these performers to his laboratory and had them participate in a range of tests measuring their speed of movement and accuracy of finger motion. But Jastrow's results revealed little out of the ordinary. He demonstrated magic has little to do with fast movements. Instead, magicians use a range of _____weapons to fool their audiences. The technique of suggestion, which captures people's minds, plays a key role in the process. In the same way that people can be made to believe that they once went on a non-existent trip in a hot-air balloon, so magicians have to be able to manipulate people's perception of performance.

① ethical ② political ③ physical
④ economic ⑤ psychological

2. Kate and her classmate, Jamie, were the youngest of the winning teams in the Stop Racism National Video Competition in 1998. Their submission depicted children playing at a nursery accompanied by messages such as "Everyone is the same; all these babies are beautiful." Kate says the point of their video was to show that human beings are not genetically coded with racist attitudes. "_____," she explains. And she adds, "The children at the nursery don't say, 'I am not playing with you since you're black,' because they don't really know about racism." According to Kate, education begins with the younger generation. And if they grow up experiencing all sorts of cultures, they have less bias.

① Cultures are very important for winning the competition
② Different skin colors give a bad impression to babies
③ Babies should be taken care of in domestic surroundings
④ Children pick up racist ideas from their surroundings as they grow up
⑤Educational institutes are free from racism due to the government's efforts

3. In this modern world, people are not used to living with discomfort. We expect immediate results and satisfaction. We want answers faster than they can be delivered. There is twenty-four-hour repair and round-the-clock shopping. If we are hungry, there is always food available, from microwave dinners to all-night grocery stores and restaurants. People no longer know how to wait, or even what waiting means. It is nice to have what you want when you want it, but the ability to delay satisfaction is important. is clearly an important virtue, yet so many people stand in front of their microwaves thinking "Hurry up!"

① Ambition ② Patience ③ Honesty
④ Modesty ⑤ Diligence

4. When we have made an error, as for example in adding up a column of figures, we have a tendency to repeat it again and again. This phenomenon is known as the persistent error. The

same thing happens when we try to solve a problem; each time our thoughts take a certain course, that course is more likely to be followed the next time. The reason that we keep making the same error repeatedly is that associations form between the ideas in the chain of thoughts and become firmer each time they are used, until finally the connections are so well established that _____. Thus, once we have adopted an unprofitable line of thought, it is harder to adopt a profitable line.

① the chain is very difficult to break
② persistent problems are solved automatically
③ the ideas lose their associations with one another
④ those connections become weaker as time goes by
⑤ the phenomenon will lead to a profitable line of thought

5. About three percent of the weight of sea water is accounted for by salt. The salt content of the ocean is being continually added to by runoff from the land, but it is not increasing. Yet salt cannot leave the sea by evaporation because the water vapor leaves it behind. So for many years it was a mystery as to why the sea should _____, given the rate of runoff, and particularly if the world was supposed to be hundreds of millions of years old. The riddle was not solved until the 1970s, when the deep ocean openings were discovered. Sea water is swallowed up by these cracks in the ocean bed; when it reemerges, millions of years later, as steam from volcanoes, the salt has been filtered out of it on its passage through the rocks.

① be so deep and large
② create so much runoff
③ have so little salt in it
④ have so many volcanoes
⑤ keep its temperature so steady

6. Furniture is among the oldest engineering structures designed to carry a rather well-defined load under rather well-defined circumstances. We are not surprised that furniture used beyond its intended purpose is broken, and we readily blame the child who abuses the furniture rather than the designer of the furniture or the furniture itself when it is abused. Thus a chair must support a person in a sitting position, but it might not be expected to survive a fall from a tall building. A child's bed might be expected to support a sleeping child, but it would not necessarily be considered badly designed if it collapsed under the child's wild use of it as a trampoline. The arms and legs of chairs, the heads and feet of beds, just like those of the people whom they serve, cannot be expected to be _____ without limit.
*trampoline: (도약용) 놀이 기구

① comfortable ② expensive ③ beautiful
④ strong ⑤ heavy

7. In Chinese food, the idea is that it should be boiling hot, because that is crucial to its flavor, embodied in the phrase wok hei, which means the 'breath' or essence of the combination of tastes added by a hot wok. In 2005 Belgian researchers at Leuven University confirmed just how the link between temperature and taste works. They identified microscopic channels in our taste buds, which seem to respond differently at different temperatures. Apparently, the higher the temperature, the more intense the flavor. This is why

_____, which is why ice cream makers add stacks of sugar—as you can tell all too clearly when ice cream melts. In a similar way, some bitter tastes, like tea, taste better when hot because they are more intense.
* wok: 중국 요리용 냄비

① ice cream tastes better when tea flavors are added
② ice cream does not taste that sweet straight from the fridge
③ they serve ice cream for dessert in Chinese restaurants
④ it is not recommended to eat ice cream while drinking hot tea
⑤ ice cream tastes sweeter especially in the winter time

8. The ultimate power is the power to get people to do as you wish. When you can do this without having to force people or hurt them, when they willingly grant you what you desire, then your power is untouchable. The best way to achieve this position is to create a relationship of dependence. The master requires your services; he is weak, or unable to function without you; you have involved yourself in his work so deeply that doing away with you would bring him great difficulty, or at least would mean valuable time lost in training another to replace you. Once such a relationship is established, you have the upper hand to make the master do as you wish. It is the classic case of the servant of the king who actually _____ the king.

① controls ② avoids ③ admires
④ rescues ⑤ entertains

9. So far as you are wholly concentrated on bringing about a certain result, clearly the quicker and easier it is brought about the better. Your resolve to secure a sufficiency of food for yourself and your family will induce you to spend weary days in tilling the ground and tending livestock; but if Nature provided food and meat in abundance ready for the table, you would thank Nature for sparing you much labor and consider yourself so much the better off. An executed purpose, in short, is a transaction in which the time and energy spent on the execution are balanced against the resulting assets, and the ideal case is one in which . Purpose, then, justifies the efforts it exacts only conditionally, by their fruits.

① demand exceeds supply, resulting in greater returns
② life becomes fruitful with our endless pursuit of dreams
③ the time and energy are limitless and assets are abundant
④ Nature does not reward those who do not exert efforts
⑤ the former approximates to zero and the latter to infinity

10. Journeys are the midwives of thought. Few places are more conducive to internal conversations than a moving plane, ship, or train. There is an almost peculiar correlation between what is in front of our eyes and the thoughts we are able to have in our heads: large thoughts at times requiring large views, new thoughts new places. Introspective reflections which are liable to stall are helped along by the flow of the landscape. The mind when thinking is all it is supposed to do. The task can be as paralyzing as having to tell a joke or mimic an accent on demand. Thinking improves when parts of the mind are given other tasks, are charged with listening to music or following a line of trees.

① may be reluctant to think properly
② may focus better on future thoughts
③ can become confused by multitasking
④ is likely to be paralyzed by fear of new tasks
⑤ can be distracted from what is before the eyes

11. Science is making the future, and nations are busy making future scientists. The more science that emerges from this investment, the greater the need for us to follow the gist of the science with sufficient understanding. In other words, if we the ordinary people are to keep pace with science, we need more science writers, and more science writing that is clear, wise and eloquent, and that demands to be read. People often feel excluded from science, convinced that it takes an advanced degree to understand what scientists do. As a result, they defensively shrug off the whole business as an exclusive realm of little relevance to their lives. One of the surest cures for scientific is great scientific literature, writing that does not merely translate technical terms into plain English or explain complicated ideas simply.

① intolerance ② immorality ③ illiteracy
④ irregularity ⑤ manipulation

12. Consumers of different age groups obviously have very different needs and wants. Although people who belong to the same age group differ in many other ways, they do tend to share a set of values and common cultural experiences that they carry throughout life. In some cases, marketers initially develop a product to attract one age group and then try to . That is what the high-octane energy drink Reddox does. The company aggressively introduced it in bars, nightclubs, and gyms to the product's core audience of young people. Over time, it became popular in other contexts, and the company began to sponsor the *Phrase Blank*A European Tour to expand its reach to older golfers. It also hands out free cans to commuters, cab drivers, and car rental agencies to promote the drink as a way to stay alert on the road.

① raise its retail price
② broaden its appeal later on
③ upgrade it for other age groups
④ increase demand by limiting supply
⑤ create a positive image via the mass media

13. Interestingly, people are more overconfident when they feel like they have control of the outcome — even when this is clearly not the case. For example, it is documented that if people are asked to bet on whether a coin toss is heads or tails, most bet larger amounts if the coin is yet to be tossed. If the coin is tossed and the outcome is concealed, people will offer lower amounts when asked for bets. People act as if _____. In this case, control of the outcome is clearly an illusion. This perception occurs in investing, as well. Even without information, people believe the stocks they own will perform better than stocks they do not own. However, ownership of a stock only gives the illusion of having control of the performance of the stock.

① the amount of the bet will influence the outcome

② their involvement will somehow affect the outcome of the toss
③ there is a parallel between a coin toss and stock investments
④ their illusion will not disappear even after the coin is tossed
⑤ they can predict the outcome with credible information

14. The identical claim, expressed in two social contexts, may have different qualifiers. When talking among friends, you might say, “Lucé is the world’s finest restaurant.” When speaking to a group of French chefs, you might find yourself saying, “Lucé is an excellent restaurant, comparable to some of the best in France.” Why did you say it differently? Perhaps because you expected a different critical scrutiny in the two groups. Maybe because your was strong enough for friends but not as strong among the most knowledgeable. In each instance, you communicated the extent to which you wanted to qualify your claim, to guard yourself by restricting the extent to which you are willing to be held accountable for the claim.

① desire to win friendship
② confidence in the claim
③ appetite for French cuisine
④ support for others’ opinions
⑤ suspicion of popular beliefs

15. By likening the eye to a camera, elementary biology textbooks help to produce a misleading impression of what perception entails. Only in terms of the physics of image formation do the eye and camera have anything in common. Both eye and camera have a lens that focuses light rays from the outside world into an image, and both have a means of adjusting the focus and brightness of that image. Both eye and camera have a light-sensitive layer onto which the image is cast (the retina and film, respectively). However, image formation is only the first step towards seeing. obscure the much more fundamental difference between the two, which is that the camera merely records an image, whereas the visual system interprets it.

① Apparent differences in the focusing power of a lens
② Superficial analogies between the eye and a camera
③ Contrasts in light adaptation between the retina and film
④ Misunderstandings of image formation in the eye and a camera
⑤ Close relationships between image formation and interpretation

16. Imagine a child playing on the beach below a cliff. He finds a cave, and full of excitement, goes in. Suddenly fear seizes him. In the deep dark of the cave, he cannot see the way ahead. What is frightening him is the sense of the unknown stretching into the black distance. Worries can be like this. Our anxiety is not about something specific, but more of a sense that unknown and uncertain possibilities may be out of sight far ahead. We can stop these worries from growing. A powerful torch or flashlight could have shown the child the limits of the cave. We can by asking: “What is the worst that can happen?” More often than not, the worst that we fear is much less terrible than our vague, unarticulated fear. Once we know the worst, we can face it directly and work out more sensibly what to do.

① hide our fears

② increase our uncertainties
③ place limits on our worries
④ share specific worries with others
⑤ differentiate reality from the ideal

17. When we behave irrationally, our behavior usually seems reasonable to us. When challenged, the mind says (to itself), "Why are these people giving me a hard time? I'm just doing what makes sense. Any reasonable person would see that!" In short, we naturally think that our thinking is fully justified. As far as we can tell, we are only doing what is right and proper and reasonable. Any fleeting thoughts suggesting that we might be at fault typically are by more powerful self-justifying thoughts: "I don't mean any harm. I'm just! I'm fair! It's the others who are wrong!" It is important to recognize this nature of the human mind as its natural state. In other words, humans don't have to learn self-justifying, self-serving, self-deceptive thinking and behavior. These patterns are innate in every one of us.

① spread ② unveiled ③ fortified
④ overcome ⑤ authorized

18. It is a common misconception among many musicians and non-musicians alike that . This is not surprising as it is natural to associate music with the sounds that create the melody, rather than with the quiet spaces between the notes. Because rests are silent, people often misinterpret these empty spaces as unimportant. But, imagine what would happen if a song was made up of only notes, and no rests. Aside from the fact that the "rests would be history" (pun intended), there would be a wall of sound with no reference point or discernible backbone to the music. This is because the spaces between the sounds provide a baseline and contrast for the piece, and give music structure and texture. In fact, it is a common saying among experienced musicians that a full measure of rest can hold more music than a full measure of blistering notes.

① notes are more important than rests
② rests provide a direct reference point to music
③ silence is no less meaningful than sound in music
④ melody is nothing more than a collection of sounds
⑤ structure and texture are the most crucial aspects of music

Test Format A

1. Instead, magicians use a range of _____ weapons to fool their audiences.

① ethical ② political ③ physical
④ economic ⑤ psychological

2. Kate says the point of their video was to show that human beings are not genetically coded with racist attitudes. "_____," she explains. And she adds, "The children at the nursery

don't say, 'I am not playing with you since you're black.' because they don't really know about racism."

① Cultures are very important for winning the competition
② Different skin colors give a bad impression to babies
③ Babies should be taken care of in domestic surroundings
④ Children pick up racist ideas from their surroundings as they grow up
⑤ Educational institutes are free from racism due to the government's efforts

3. In this modern world, people are not used to living with discomfort. We expect immediate results and satisfaction.

People no longer know how to wait, or even what waiting means. It is nice to have what you want when you want it, but the ability to delay satisfaction is important. _____ is clearly an important virtue, yet so many people stand in front of their microwaves thinking "Hurry up!"

① Ambition ② Patience ③ Honesty

④ Modesty ⑤ Diligence

4. When we have made an error, as for example in adding up a column of figures, we have a tendency to repeat it again and again. This phenomenon is known as the persistent error. The same thing happens when we try to solve a problem; each time our thoughts take a certain course, that course is more likely to be followed the next time. The reason that we keep making the same error repeatedly is that associations form between the ideas in the chain of thoughts and become firmer each time they are used, until finally the connections are so well established that _____. Thus, once we have adopted an unprofitable line of thought, it is harder to adopt a profitable line.

① the chain is very difficult to break
② persistent problems are solved automatically
③ the ideas lose their associations with one another
④ those connections become weaker as time goes by
⑤ the phenomenon will lead to a profitable line of thought

5. So for many years it was a mystery as to why the sea should _____, given the rate of runoff, and particularly if the world was supposed to be hundreds of millions of years old.

① be so deep and large
② create so much runoff
③ have so little salt in it
④ have so many volcanoes
⑤ keep its temperature so steady

6. A child's bed might be expected to support a sleeping child, but it would not necessarily be considered badly designed if it collapsed under the child's wild use of it as a trampoline. The

arms and legs of chairs, the heads and feet of beds, just like those of the people whom they serve, cannot be expected to be _____ without limit.
*trampoline: (도약용) 놀이 기구

① comfortable ② expensive ③ beautiful
④ strong ⑤ heavy

7. In Chinese food, the idea is that it should be boiling hot, because that is crucial to its flavor, embodied in the phrase wok hei, which means the 'breath' or essence of the combination of tastes added by a hot wok.

Apparently, the higher the temperature, the more intense the flavor. This is why _____, which is why ice cream makers add stacks of sugar—as you can tell all too clearly when ice cream melts. In a similar way, some bitter tastes, like tea, taste better when hot because they are more intense.
* wok: 중국 요리용 냄비

① ice cream tastes better when tea flavors are added
② ice cream does not taste that sweet straight from the fridge
③ they serve ice cream for dessert in Chinese restaurants
④ it is not recommended to eat ice cream while drinking hot tea
⑤ ice cream tastes sweeter especially in the winter time

8. The ultimate power is the power to get people to do as you wish. When you can do this without having to force people or hurt them, when they willingly grant you what you desire, then your power is untouchable. The best way to achieve this position is to create a relationship of dependence. The master requires your services; he is weak, or unable to function without you; you have involved yourself in his work so deeply that doing away with you would bring him great difficulty, or at least would mean valuable time lost in training another to replace you. Once such a relationship is established, you have the upper hand to make the master do as you wish. It is the classic case of the servant of the king who actually _____ the king.
① controls ② avoids ③ admires
④ rescues ⑤ entertains

9. An executed purpose, in short, is a transaction in which the time and energy spent on the execution are balanced against the resulting assets, and the ideal case is one in which _____.

① demand exceeds supply, resulting in greater returns
② life becomes fruitful with our endless pursuit of dreams
③ the time and energy are limitless and assets are abundant
④ Nature does not reward those who do not exert efforts
⑤ the former approximates to zero and the latter to infinity

10. Introspective reflections which are liable to stall are helped along by the flow of the landscape. The mind _____ when thinking is all it is supposed to do. The task can be as paralyzing as having to tell a joke or mimic an accent on demand.

① may be reluctant to think properly
② may focus better on future thoughts
③ can become confused by multitasking
④ is likely to be paralyzed by fear of new tasks
⑤ can be distracted from what is before the eyes

11. Science is making the future, and nations are busy making future scientists.

As a result, they defensively shrug off the whole business as an exclusive realm of little relevance to their lives. One of the surest cures for scientific _____ is great scientific literature, writing that does not merely translate technical terms into plain English or explain complicated ideas simply.

① intolerance ② immorality ③ illiteracy
④ irregularity ⑤ manipulation

12. Consumers of different age groups obviously have very different needs and wants. Although people who belong to the same age group differ in many other ways, they do tend to share a set of values and common cultural experiences that they carry throughout life. In some cases, marketers initially develop a product to attract one age group and then try to _____. That is what the high-octane energy drink Reddox does. The company aggressively introduced it in bars, nightclubs, and gyms to the product's core audience of young people. Over time, it became popular in other contexts, and the company began to sponsor the *Phrase Blank*A European Tour to expand its reach to older golfers. It also hands out free cans to commuters, cab drivers, and car rental agencies to promote the drink as a way to stay alert on the road.
① raise its retail price
② broaden its appeal later on
③ upgrade it for other age groups
④ increase demand by limiting supply
⑤ create a positive image via the mass media

13. People act as if _____.

① the amount of the bet will influence the outcome
② their involvement will somehow affect the outcome of the toss
③ there is a parallel between a coin toss and stock investments
④ their illusion will not disappear even after the coin is tossed
⑤ they can predict the outcome with credible information

14. Perhaps because you expected a different critical scrutiny in the two groups. Maybe because your _____ was strong enough for friends but not as strong among the most knowledgeable. In each instance, you communicated the extent to which you wanted to qualify

your claim, to guard yourself by restricting the extent to which you are willing to be held accountable for the claim.

① desire to win friendship
② confidence in the claim
③ appetite for French cuisine
④ support for others' opinions
⑤ suspicion of popular beliefs

15. By likening the eye to a camera, elementary biology textbooks help to produce a misleading impression of what perception entails.

However, image formation is only the first step towards seeing. _____ obscure the much more fundamental difference between the two, which is that the camera merely records an image, whereas the visual system interprets it.

① Apparent differences in the focusing power of a lens
② Superficial analogies between the eye and a camera
③ Contrasts in light adaptation between the retina and film
④ Misunderstandings of image formation in the eye and a camera
⑤ Close relationships between image formation and interpretation

16. Imagine a child playing on the beach below a cliff. He finds a cave, and full of excitement, goes in. Suddenly fear seizes him. In the deep dark of the cave, he cannot see the way ahead. What is frightening him is the sense of the unknown stretching into the black distance. Worries can be like this. Our anxiety is not about something specific, but more of a sense that unknown and uncertain possibilities may be out of sight far ahead. We can stop these worries from growing. A powerful torch or flashlight could have shown the child the limits of the cave. We can _____ by asking: "What is the worst that can happen?" More often than not, the worst that we fear is much less terrible than our vague, unarticulated fear. Once we know the worst, we can face it directly and work out more sensibly what to do.

① hide our fears
② increase our uncertainties
③ place limits on our worries
④ share specific worries with others
⑤ differentiate reality from the ideal

17. Any fleeting thoughts suggesting that we might be at fault typically are _____ by more powerful self-justifying thoughts: "I don't mean any harm. I'm just! I'm fair! It's the others who are wrong!"
① spread ② unveiled ③ fortified
④ overcome ⑤ authorized

18. It is a common misconception among many musicians and non-musicians alike that _____. This is not surprising as it is natural to associate music with the sounds that create the melody, rather than with the quiet spaces between the notes.

① notes are more important than rests
② rests provide a direct reference point to music
③ silence is no less meaningful than sound in music
④ melody is nothing more than a collection of sounds
⑤ structure and texture are the most crucial aspects of music

Questionnaire (Korean)

19. 내가 푼 문제지 유형 (시험지 오른쪽 상단에 있습니다.)	① 유형 A	② 유형 B	③ 유형 C	④ 유형 D	
20. 나의 대학수학능력평가 영어영역 등급 (20 번과 21 번 통틀어 하나만 선택하세요.)	① 1 등급	② 2 등급	③ 3 등급	④ 4 등급	⑤ 5 등급
21. 나의 대학수학능력평가 영어영역 등급 (20 번과 21 번 통틀어 하나만 선택하세요.)	① 6 등급	② 7 등급	③ 8 등급	④ 9 등급	
22. 나는 영어권국가에 1 년 이상 거주하거나 외국에서 영어를 사용하는 국제학교를 1 년 이상 다닌 경험이 있다. (‘아니오’라고 답한 경우 바로 24 번으로)	① 예	② 아니오			
23. 22 번에 ‘예’라고 답한 경우 외국에 거주한 기간	① 1 년	② 1 년~3 년	③ 3 년~5 년	④ 5 년이상	
24. 오늘 풀어본 문항 (1 번부터 18 번까지) 중 지문을 미리 읽어보거나 공부한 적이 있다.	① 예	② 아니오			
25. 24 번에 ‘예’라고 답한 경우, 내가 미리 읽어보거나 공부한 적이 있는 문항은 다음과 같다.	① 1 번	② 2 번	③ 3 번	④ 4 번	⑤ 5 번

26. 24 번에 ‘예’라고 답한 경우, 내가 미리 읽어보거나 공부한 적이 있는 문항은 다음과 같다.	① 6 번	② 7 번	③ 8 번	④ 9 번	⑤ 10 번
27. 24 번에 ‘예’라고 답한 경우, 내가 미리 읽어보거나 공부한 적이 있는 문항은 다음과 같다.	① 11 번	② 12 번	③ 13 번	④ 14 번	⑤ 15 번
28. 24 번에 ‘예’라고 답한 경우, 내가 미리 읽어보거나 공부한 적이 있는 문항은 다음과 같다.	① 16 번	② 17 번	③ 18 번		

전혀 아니다 ←→ 매우 그렇다

29. 나는 수능 및 모의고사 영어영역의 문제를 풀 때 전반적으로 글을 다 읽고 문제를 푼다.	①	②	③	④	⑤
30. 나는 빈칸추론문항을 풀 때 글을 다 읽고 문제를 푼다.	①	②	③	④	⑤
31. 나는 수능 및 모의고사 영어영역의 문제를 풀 때 글보다 문제와 선택지를 먼저 보고 푼다.	①	②	③	④	⑤
32. 나는 빈칸추론문항을 풀 때 글보다 문제와 선택지를 먼저 보고 푼다.	①	②	③	④	⑤

33. 빈칸추론문항을 풀 때 가장 먼저 보는 부분은 다음과 같다.

① 글의 첫 문장
② 글의 마지막 문장
③ 빈칸이 있는 문장의 바로 앞 문장
④ 빈칸이 있는 문장
⑤ 빈칸이 있는 문장의 바로 다음 문장

34. 빈칸추론 문항을 풀 때 글을 처음부터 읽지 않는다면 그 이유는 다음과 같다.

① 시간이 부족해서
② 난이도가 높아서 (어려워서)
③ 다른 곳부터 읽으라고 배웠기 때문에

Appendix B: Post-hoc result of multiple comparisons (Bonferroni)

	The English proficiency level		Mean Difference	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Q1	Level 4	Level 1	.78746*	0.16936	0*	0.3374	1.2376
		Level 2	.52920*	0.17626	0.018*	0.0608	0.9976
		Level 3	.56322*	0.16733	0.005*	0.1185	1.0079
	Level 3	Level 1	0.22424	0.19553	1	-0.2954	0.7439
		Level 2	-0.03401	0.20153	1	-0.5696	0.5016
		Level 4	-.56322*	0.16733	0.005*	-1.0079	-0.1185
	Level 2	Level 1	0.25826	0.20322	1	-0.2818	0.7983
		Level 3	0.03401	0.20153	1	-0.5016	0.5696
		Level 4	-.52920*	0.17626	0.018*	-0.9976	-0.0608
	Level 1	Level 2	-0.25826	0.20322	1	-0.7983	0.2818
		Level 3	-0.22424	0.19553	1	-0.7439	0.2954
		Level 4	-.78746*	0.16936	0*	-1.2376	-0.3374
Q2	Level 4	Level 1	1.12837*	0.17757	0*	0.6565	1.6003
		Level 2	.75880*	0.1848	0*	0.2677	1.2499
		Level 3	.70796*	0.17544	0*	0.2417	1.1742
	Level 3	Level 1	0.42041	0.20501	0.247	-0.1244	0.9652
		Level 2	0.05084	0.2113	1	-0.5107	0.6124
		Level 4	-.70796*	0.17544	0*	-1.1742	-0.2417
	Level 2	Level 1	0.36957	0.21307	0.504	-0.1967	0.9358
		Level 3	-0.05084	0.2113	1	-0.6124	0.5107
		Level 4	-.75880*	0.1848	0*	-1.2499	-0.2677
	Level 1	Level 2	-0.36957	0.21307	0.504	-0.9358	0.1967
		Level 3	-0.42041	0.20501	0.247	-0.9652	0.1244
		Level 4	-1.12837*	0.17757	0*	-1.6003	-0.6565
Q3	Level 4	Level 1	.70627*	0.18878	0.001*	0.2046	1.208

		Level 2	0.4803	0.19647	0.091	-0.0418	1.0024
		Level 3	0.24471	0.18652	1	-0.251	0.7404
	Level 3	Level 1	0.46156	0.21795	0.211	-0.1177	1.0408
		Level 2	0.23559	0.22464	1	-0.3614	0.8326
		Level 4	-0.24471	0.18652	1	-0.7404	0.251
	Level 2	Level 1	0.22597	0.22652	1	-0.376	0.828
		Level 3	-0.23559	0.22464	1	-0.8326	0.3614
		Level 4	-0.4803	0.19647	0.091	-1.0024	0.0418
	Level 1	Level 2	-0.22597	0.22652	1	-0.828	0.376
		Level 3	-0.46156	0.21795	0.211	-1.0408	0.1177
		Level 4	-.70627*	0.18878	0.001*	-1.208	-0.2046
Q4	Level 4	Level 1	0.49723	0.18757	0.051*	-0.0013	0.9957
		Level 2	0.27754	0.1966	0.955	-0.245	0.8001
		Level 3	0.10648	0.18533	1	-0.3861	0.599
	Level 3	Level 1	0.39075	0.21625	0.431	-0.184	0.9655
		Level 2	0.17105	0.22413	1	-0.4246	0.7667
		Level 4	-0.10648	0.18533	1	-0.599	0.3861
	Level 2	Level 1	0.2197	0.22599	1	-0.3809	0.8203
		Level 3	-0.17105	0.22413	1	-0.7667	0.4246
		Level 4	-0.27754	0.1966	0.955	-0.8001	0.245
	Level 1	Level 2	-0.2197	0.22599	1	-0.8203	0.3809
		Level 3	-0.39075	0.21625	0.431	-0.9655	0.184
		Level 4	-0.49723	0.18757	0.051*	-0.9957	0.0013

국 문 초 록

한국 EFL 고등학교 학습자의 대학수학능력시험 빈칸추론문항 수행:

지문 정보의 양, 빈칸의 종류, 모의고사 영어 등급을 중심으로

김 지 은

외국어교육과 영어전공

서울대학교 대학원

본 연구는 모의고사 영어 등급, 지문 정보의 양, 빈칸의 종류에 따른 한국 EFL 고등학생 학습자들의 빈칸추론문항 수행을 탐구하였다. 모의고사 등급에 따라 네 그룹으로 나누어진 279 명의 고등학교 2 학년 학생들이 대학수학능력시험과 평가원 모의고사 기출에서 선택한 18 개의 빈칸 추론문항에 응답하였다. 문항들을 각각 다른 지문 정보의 양을 포함하도록 빈칸을 포함한 문장 (Sentence-with-a-blank), 문장을 포함한 문장과 그 앞뒤 문장 (Left-and-right Passage), 빈 칸을 포함한 문장과 그 앞뒤 문장 그리고 글의 가장 처음과 마지막 문장 (First-and-final Passage), 글 전체 (Original Passage)로 구성된 네 가지 양식으로 나누었다. 또한 빈칸에 들어갈 말의 종류에 따라 문항들을 한 단어, 구, 절의 세 가지 빈칸유형으로 나누었다.

이 연구는 학습자들의 빈칸추론문항 수행에 대한 두 가지 발견점을 시사한다. 첫 번째, 글 전체를 읽는 것은 모든 학습자들에게 추가적인 도움이

되지 않았지만 중요한 세 문장만을 읽었을 때 모든 학습자들은 가장 높은 수행을 보였다. 학습자들은 모의고사 등급에 따라 반대의 수행을 보였는데 가장 낮은 수준의 학습자가 가장 긴 글에서 가장 낮은 정답을 맞출 확률을 보인 반면, 가장 높은 수준의 학습자들은 가장 긴 글에서 가장 높은 정답을 맞출 확률 보였다. 두 번째로, 낮은 수준의 학습자들은 빈칸에 들어갈 말의 종류에 따라 어려움을 느끼는 정도가 다른 반면 높은 수준의 학습자들은 빈칸종류의 효과에 덜 민감하다. 빈칸에 절이 들어가는 문항은 한 단어가 들어가는 문항보다 더 많은 읽기를 요구하기 때문에 더 어렵고 구가 들어가는 문항은 문맥에 더 영향을 많이 받기 때문에 가장 어렵다. 가장 낮은 수준의 학습자들은 빈칸에 들어갈 정보를 찾기 위해 긴 선택지를 읽고 완전히 이해하는데 실패하기 쉽다.

결론적으로 고등학교 학생들은 빈칸추론문항에서 주요한 세 문장을 읽었을 때에도 지문 전체를 읽었을 때만큼의 정답을 맞출 확률을 보일 수 있다. 하지만 그 주요 세 문장과 지문 전체를 읽는 것이 같은 읽기 능력을 측정하는지는 재고해보아야 한다.

주요어: 빈칸추론문항, 시험 타당도, 객관식 문항, 읽기 평가, 제 2 언어 읽기

학 번: 2017-27271